第5章　推定の規準と方法

まえがき　4章では，観測値の1次と2次の積率の指定された構造だけを用いて，母数の推定を考えた．この章では，観測値の分布関数が，ある分布関数の族$\mathscr{F}$に属するということが指定されているというより一般的な状況を考える．そのとき，より広い推定問題は，観測されたデータつまり実現した観測値に基づいて，$\mathscr{F}$に属する1つの分布関数を選択するという問題である．いくつかの問題では，分布関数の特別な特性（母数あるいはその関数）にだけ興味があり，その場合には，母数だけを推定することが必要であって，分布関数全部を推定する必要は必ずしもない．

どちらの場合でも，必要なものは観測値の関数であり，それは**推定量**と呼ばれ，実現された観測値にたいするその値は，未知の量（分布関数あるいは母数）の**推定値**である．推定量を決定するためには，その良さを判断する規準を必要とする．明らかに，その規準は推定値を得る目的に依存する．したがって，すべての推定量の良さを判断することのできるただ1つの規準があるものでもなく，どんな場合についても適当であるような母数のただ1つの推定量があるものでもない．この章では，さまざまな推定の規準について議論し，推定量を得るための適当な手順を与えることを試みる．

推定の方法と推定値の不確かさの表現については議論が数多い．それらについては，後で少し論じる．

章末の問題は，推定についての多くの結果を含んでいる．それらは，推定の理論と方法についての補足として読むことができる．1つの数値例が，最尤推定の得点による方法を示すために，章末にあげられている．

5a　最小分散不偏推定

5a.1　最小分散規準

4章では，母数の関数を推定するために，最小2乗法を導入した．そこでは，任意の与えられた母数の関数にたいして，最小2乗推定量が線形不偏推定量のなかで最小分散をもつことが示された．最小2乗推定量が，すべての不偏推定量のなかでどのような推定量で

あるかは，観測値の分布について1次と2次のモーメントだけしか指定されなかったので，研究することができなかった．したがって，観測値の分布がより狭いクラスに制約された場合に，すべての不偏推定量のなかの最小分散推定量を得る一般的な方法を調べることは興味深い．

直観的には，母数 θ の推定量というとき，それは何らかの意味で真の値に最も近い，観測値 $(x_1, \cdots, x_n)$ の関数 T を意味している．推定の規準を設けるときには，推定量の，母数の真値との近さの尺度を与え，また推定量のクラスに適当な制約を課そうとする．制約されたクラスのなかで，最適な推定量は，近さの尺度を最小にすることにより決定される．$T = c$（定数）という推定量は，母数の未知の値が c であるときには，どのような意味においても最良である．このような無意味な推定量を避けるために，推定量のクラスに何らかの制約を置くことが必要であろう．

すべての場合に有用であるような推定の規準がただ1組である必要はない．統計学の文献で用いられているいく組かの規準について調べ，各々の長所，欠点について論ずることにしよう．

もちろん次のような関数 T が存在するならば理想的であろう．つまり，任意の他の統計量 T' と比較したとき，関係式

$$P(\theta - \lambda_1 < T < \theta + \lambda_2) \geqslant P(\theta - \lambda_1 < T' < \theta + \lambda_2) \qquad (5a.1.1)$$

が，選ばれた区間 $(0, \lambda)$ のすべての λ_1, λ_2 とすべての θ について成り立つ場合である．条件 (5a.1.1) がすべての λ についてみたされるならば，1つの必要条件は

$$E(T - \theta)^2 \leqslant E(T' - \theta)^2 \qquad (5a.1.2)$$

である．ここで E は期待値をあらわす．つまり，T の平均2乗誤差 (m.s.e.) が最小である．さらに，推定量が不偏でなければならないとするならば，

$$V(T) \leqslant V(T') \qquad (5a.1.3)$$

となる．ここで V は分散をあらわす．

集中度が最も高いという規準 (5a.1.1)，あるいは，平均2乗誤差が最小という規準 (5a.1.2) をみたす推定量は一般には存在しない．しかしながら，推定量が不偏でなければならないという制約を置くと，最小分散をもつという意味で最適な推定量を，いくつかの問題では，得ることができる．不偏性の条件は，より小さな m.s.e. をもつかもしれない多くの偏りのある推定量が推定関数としての資格を失なう場合には，特に，興味のない条件であるかもしれない．また，不偏推定量が存在しない場合や，不偏推定量が存在しても最小分散推定量が存在しない場合もある［章末の問題 **12** を参照］．

$N(\mu, \sigma^2)$ からの n 個の観測値 $x_1, \cdots, x_n$ から計算される統計量

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

を考えよう．$E(s^2) = \sigma^2$，つまり s^2 は σ^2 の不偏推定量であることは知られている．$(n-1)s^2/\sigma^2 \sim \chi^2(n-1)$ であるから，

$$V\left[\frac{(n-1)s^2}{\sigma^2}\right] = V[\chi^2(n-1)] = 2(n-1)$$

なので，$V(s^2) = 2\sigma^4/(n-1)$ となる．ところで，

$$E(cs^2 - \sigma^2)^2 = E[c(s^2 - \sigma^2) - \sigma^2(1-c)]^2 = \left\{\frac{2c^2}{n-1} + (1-c)^2\right\}\sigma^4$$

は，$c = (n-1)/(n+1)$ のとき最小となり，その最小値は

$$\frac{2\sigma^4}{n+1} < \frac{2\sigma^4}{n-1}$$

である．したがって，すべての n と σ にたいして

$$E\left\{\frac{\sum (x_i - \bar{x})^2}{n+1} - \sigma^2\right\}^2 < E\left\{\frac{\sum (x_i - \bar{x})^2}{n-1} - \sigma^2\right\}^2$$

となる．推定量 $\sum(x_i - \bar{x})^2/(n+1)$ は σ^2 の不偏推定量ではないが，m.s.e. の規準を用いれば，s^2 よりもよい推定量である．それでは，どちらの推定量が選ばれるべきだろうか？　明らかに，その答は推定値を得る目的に依存する．

r を，未知の成功の確率 π をもつ n 回のベルヌイ試行における成功の回数としよう．そのとき，$E(f) = 1/\pi$ となるような r の関数 f は存在しない．つまり，この場合には不偏推定ができないのである．

最小分散不偏推定の最も重要な役割は，母数の独立な推定量のもつ情報の併合にある．$T_1, T_2, \cdots$ を θ の推定量とし，

$$E(T_i) = \theta, \qquad V(T_i) = \sigma_i^2 < \sigma^2$$

とする．さらに，T_i の分布については何も指定されていないものとする．$\bar{T}_k = (T_1 + \cdots + T_k)/k$ とおけば，

$$E(\bar{T}_k) = \theta, \qquad V(\bar{T}_k) = \frac{\sum \sigma_i^2}{k^2} < \frac{\sigma^2}{k}$$

となるので，$k \to \infty$ のとき $V(\bar{T}_k) \to 0$ となる．つまり，$k \to \infty$ のとき $\bar{T}_k$ は真の値 θ にどんどん近づいていく．ところが，もしも T_i が不偏推定量ではなくて，未知の偏り β をもつとすると，$\bar{T}_k$ は，真の値 θ の代わりに誤った値 $\theta + \beta$ に近づいていく．

個々の推定量の偏りは，（標準）誤差と比較して小さいならば重大ではないかもしれない．しかし，いくつかのそのような推定量を上に示したように併合したときには，$k \to \infty$ のとき分散は0に近づくにもかかわらず，偏りはそのまま残ってしまう．ある段階を過ぎると，偏りは標準誤差に比較して大きくなるのである．

あるいは，応答 y と因子変数 x とがあり，x のいくつかの値にたいして y の互いに独立な推定値を得ることにより，2つの変数の間の関係を調べたいものとしよう．もしも x のいろいろな値にたいする y の推定量が不偏でないならば，推定された関係には系統的な誤差がはいる可能性がある．このような場合に，推定量が不偏であることが望ましいならば，真の値への速い収束を保証するためには，より小さな分散をもつ推定量が選ばれるべきであることは明らかである．

5a.2　最小分散推定についてのいくつかの基本的な結果

$\mathscr{P}_0$ を，観測値 X の標本空間上の確率測度の集合とし，指定された母数空間 Θ の中の値をとる母数 θ（ベクトル値であるかもしれない）を指標としてもつものとしよう．多くの応用では，X は一般には n 次元ベクトルであるが，ここでの議論にたいしては，より一般的なものであってよい．U_g を，θ の特定の関数 g の，すべての不偏な推定量（Xの関数）の集まりとし，U_0 を，期待値が零である関数の全体とする．したがって，$T \in U_g$ であるための必要十分条件は，すべての $\theta \in \Theta$ にたいして $E(T|\theta) = g(\theta)$ であることであり，$f \in U_0$ であるための必要十分条件は，すべての θ にたいして $E(f|\theta) = 0$ であることである．以下で，最小分散不偏推定量 m.v.u.e.(minimum variance unbiased estimator) に関するいくつかの結果を証明しよう．

（i）推定量 $T \in U_g$ は，$V(T|\theta_0) < \infty$ とする．T が $\theta = \theta_0$ で最小分散をもつための必要十分条件は，$V(f|\theta_0) < \infty$ であるようなすべての $f \in U_0$ にたいして，$\mathrm{cov}(T, f|\theta_0) = 0$ であることである．

必要性は次のようにして容易に示される．任意の λ にたいして推定量 $T + \lambda f$ は g の不偏推定量である．もしも $\mathrm{cov}(T, f) = 0$ でなければ，区間 $[0, -2\,\mathrm{cov}(T, f)/V(f)]$ の任意の λ にたいして

$$V(T + \lambda f) = V(T) + 2\lambda\,\mathrm{cov}(T, f) + \lambda^2 V(f) \leqslant V(T) \qquad (5a.2.1)$$

となるからである．十分性を証明するために，$T' \in U_g$ を θ_0 で $V(T') < \infty$ である別の推定量としよう．そうすると $(T - T') \in U_0$ であるから，条件によって

$$E[T(T - T')|\theta_0] = 0$$

となる．したがって

$$V(T) = \mathrm{cov}(T, T') \quad \text{つまり} \quad \sqrt{V(T)} = \rho\sqrt{V(T')} \leqslant \sqrt{V(T')} \qquad (5a.2.2)$$

である．ここで ρ は T と T' の相関係数である．よって（i）が証明された［証明のなかで，すべての分散と共分散は θ_0 で計算されている］．

注意　（a）m.v.u.e. と，任意の不偏推定量との相関係数は非負である［(5a.2.2) を参照］．

（b）2つの不偏推定量が同じ最小の分散をもつとき，それらの相関係数 $\rho(\theta_0)$ は1である．つまり，2つの推定量は，θ_0 について確率零の標本の集合を除いて同じである［(5a.2.2) による］．

（c）推定量を，線形関数あるいは連続関数といったような，特別な関数の集まり G に制限すれば，必要十分条件は，$V(f|\theta_0) < \infty$ であるようなすべての $f \in U_0 \cap G$ にたいして $E(Tf|\theta_0) = 0$ であることである．

（d）推定量 T を任意の集合 H に属するものに限定しよう．そのとき，m.s.e. $E(T - \theta_0)^2$ が最小であるための必要十分条件は，すべての $W \in H$ にたいして，

$$E[(T - \theta_0)(T - W)|\theta_0] = 0$$

であることである．〔"H は任意の集合" となっているが，線形空間でなければいけない．〕

（e）T_1, T_2 が各々 $g_1(\theta), g_2(\theta)$ の m.v.u. 推定量であれば，b_1, b_2 を定数としたとき $b_1T_1 + b_2T_2$ は $b_1g_1(\theta) + b_2g_2(\theta)$ の m.v.u. 推定量である［これは（i）の直接の結果である］．

（f）T がすべての θ について最小分散をもつときには，一様 m.v.u. 推定量であるという．

例として，4章で議論した最小2乗法の理論の，観測値についての模型 $(\mathbf{Y}, \mathbf{X\beta}, \sigma^2\mathbf{I})$ を考えよう．$\mathbf{Y} \sim N(\mathbf{X\beta}, \sigma^2\mathbf{I})$ とし，$f(\mathbf{Y}) \in U_0$ とすると，

$$\int f(\mathbf{Y})e^{-(\mathbf{Y}-\mathbf{X\beta})'(\mathbf{Y}-\mathbf{X\beta})/2\sigma^2}dv \equiv 0$$

あるいは，

$$\int f(\mathbf{Y})e^{-\mathbf{Y}'\mathbf{Y}/2\sigma^2+\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y}/\sigma^2}dv \equiv 0 \qquad (5a.2.3)$$

となる．(5a.2.3) を $\boldsymbol{\beta}$ の第 i 成分 β_i について微分することにより

$$\int f(\mathbf{Y})(x_{1i}y_1 + \cdots + x_{ni}y_n)e^{-(\mathbf{Y}-\mathbf{X\beta})'(\mathbf{Y}-\mathbf{X\beta})/2\sigma^2}dv = 0 \qquad (5a.2.4)$$

を得る．このことから，

$$Q_i = x_{1i}y_1 + \cdots + x_{ni}y_n,$$

したがって，Q_i の任意の線形関数 $\lambda_1Q_1 + \cdots + \lambda_mQ_m = \boldsymbol{\lambda}'\mathbf{Q} = \boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}$（行列表示をすれば）も，その期待値の一様 m.v.e. であることがわかる．ところで，(4a.2.5) で導いた通り，任意のベクトル $\boldsymbol{\lambda}$ にたいして，$\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}$ は母数の関数の最小2乗推定量 (l.s.e.) である（4章では，l.s.e. が線形不偏推定量のなかで最小分散をもつことだけが示された）．それゆえ，正規性の仮定を設けることにより，l.s.e. が不偏推定量全体のなかで最小分散をもつという，より強い結果を証明することができたのである．

著者 Rao (1959c) は，$\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}$ の各成分が，すべての次数のモーメントが有限であるような同じ分布をしており，かつ，l. s. 推定量が不偏推定量全体のなかで最小分散をもつならば，$\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}$ の各成分は正規分布をしている，ということを示した．したがって，正規分布の新しい型の特徴づけができたことになる．

(5a. 2. 4) を β_j について微分すれば，$E(fQ_iQ_j)=0$ が得られるので，Q_iQ_j が m. v. e. であることがわかる．一方 (5a. 2. 3) を σ^2 について微分すれば，$\mathbf{Y}'\mathbf{Y}$ が m. v. e. であることもわかる．それゆえ，任意の定数 b_{ij} にたいして，$\mathbf{Y}'\mathbf{Y}-\sum\sum b_{ij}Q_iQ_j$ は m. v. e. である．ところで，適当に b_{ij} を選べば，

$$\mathbf{Y}'\mathbf{Y}-\sum\sum b_{ij}Q_iQ_j=\mathbf{Y}'\mathbf{Y}-\mathbf{Q}'\mathbf{B}\mathbf{Q}=\mathbf{Y}'\mathbf{Y}-\mathbf{Y}'\mathbf{X}\mathbf{B}\mathbf{X}'\mathbf{Y}=R_0^2$$

[ここで $\mathbf{B}=(b_{ij})$] は，最小2乗法の理論における残差平方和であり，$(n-r)\sigma^2$ の不偏推定量である．したがって，正規性の仮定をおけば，$R_0^2/(n-r)$ は不偏推定量全体のなかで，σ^2 の m. v. u. e. である．

推定量の関数形に制約をおく例として，n 個の観測値 $x_1,\cdots,x_n$ の確率分布が $P(x_1-\theta,\cdots,x_n-\theta)$ という形の場合を考えよう．推定関数を，移行性，つまり

$$f(x_1+\alpha,\cdots,x_n+\alpha)=\alpha+f(x_1,\cdots,x_n),\quad 任意の\ \alpha\ にたいして \tag{5a. 2. 5}$$

という性質をみたすものに限定しよう．関数 $\bar{x}=(x_1+\cdots+x_n)/n$ は性質 (5a. 2. 5) をみたしている．(5a. 2. 5) をみたす任意の関数を考え，$\bar{x}$ との差を h とすれば，h は差 $y_1=x_2-x_1,\cdots,y_{n-1}=x_n-x_{n-1}$ だけの関数である．y で $y_1,\cdots,y_{n-1}$ 全体をあらわすことにする．そうすると，次のことが容易にわかる．

$$E(\bar{x}\,|\,y)=\theta+e(y),$$

ここで $e(y)$ は θ を含まない．さらに

$$\begin{aligned}E[\bar{x}+h(y)-\theta]^2&=\underset{y}{E}\,\underset{\bar{x}}{E}[\bar{x}-\theta-e(y)+e(y)+h(y)]^2\\&=E[v(y)]+E[e(y)+h(y)]^2\end{aligned} \tag{5a. 2. 6}$$

ここで $v(y)=E([\bar{x}-\theta-e(y)]^2\,|\,y)$ は $h(y)$ を含まない．したがって，(5a. 2. 6) は $h(y)=-e(y)$ のときに最小となる．このように $h(y)$ を選べば，統計量 $\bar{x}-e(y)$ は，(5a. 2. 5) をみたし，不偏推定量であるばかりでなく，一様に最小の平均2乗誤差をもつ．他の例については Rao (1952a, 1952b, c) を参照せよ．

(ii) 標本が，同じ分布からの n 個の独立な観測値から成り立っているならば，最小分散不偏推定量は観測値について対称である．

$T(x_1,\cdots,x_n)$ を母数の不偏推定量とするならば，対称な関数 $\sum T(x_a,x_b,\cdots)/n!$ もまた不偏推定量であり，分散は T よりも大きくはないということが容易に示される．ここで和 $\sum$ は $x_1,\cdots,x_n$ のすべての可能な順列についてとる．

$V[T(x_1,\cdots,x_n)]=\sigma^2$ とする．そうすると，対称性により，$x_1,\cdots,x_n$ の任意の順列にたいして，$V[T(x_a,x_b,\cdots)]=\sigma^2$ である．さらに，任意の2つの順列 $x_a,x_b,\cdots$ と $x_{a'},x_{b'},\cdots$ にたいして，$\mathrm{cov}[T(x_a,x_b,\cdots),T(x_{a'},x_{b'},\cdots)]\leqslant\sigma^2$ である．したがって，

$$\begin{aligned}V\Big[\sum T(x_a,x_b,\cdots)/n!\Big]&=\frac{1}{(n!)^2}\sum V[T(x_a,x_b,\cdots)]\\&\quad+\frac{1}{(n!)^2}\sum\sum\mathrm{cov}[T(x_a,x_b,\cdots),\,T(x_{a'},x_{b'},\cdots)]\leqslant\sigma^2\frac{n!+n!(n!-1)}{(n!)^2}=\sigma^2\end{aligned}$$

となる．

結果 (ii) が重要であることは，分布の正確な形がわからないとき，母数を推定する場合に了解されるだろう．たとえば，$\bar{x}=(x_1+\cdots+x_n)/n$ は対称であり，分布の平均の不偏推定量である．この推定量は，分布の形について，1次のモーメントが存在するという以外には何も知られていないときには，ただ1つの不偏推定量である．なぜならば $\bar{x}$ のほかに，x_i について対称な不偏推定量 $\bar{x}+\delta$ があったとしよう．そのとき，δ は x_i について対称で $E[\delta]=0$ でなければならないので，特定の分布を選ぶことにより $\delta\equiv 0$ が示せるのである．同様に，$x_1,\cdots,x_n$ が i. i. d. (independent and identically distributed) なる変量ならば，

$$s^2=\frac{x_1^2+\cdots+x_n^2-n\bar{x}^2}{n-1}$$

は，x_i の分布について何も指定されていないとき，母集団分散の m. v. 不偏推定量である．

(iii) T を θ の十分統計量とする．ここで T も θ もベクトルであってもよいものとする．T_1 を任意の統計量とし，g を θ の任意の関数とすれば，

$$E[T_1-g(\theta)]^2\geqslant E[h(T)-g(\theta)]^2$$

である．ここで $h(t)=E(T_1\,|\,T=t)$ であり，θ とは無関係である．T_1 が $g(\theta)$ の不偏推定量であれば，$h(T)$ もそうである．

T は θ の十分統計量であるから，$T=t$ のもとでの T_1 の条件つき期待値は，θ と無関係である．したがって，$h(t)$ は θ と無関係である．条件つき期待値についての結果 (2b. 3. 4) を用いれば，

$$E(T_1)=\underset{T}{E}[E(T_1\,|\,T)]=E[h(T)] \tag{5a. 2. 7}$$

であるので，T_1 と $h(T)$ とは同じ期待値をもつことがわかる．さらに，(2b. 3. 7) を用いると

$$\begin{aligned}E\{h(T)[T_1-h(T)]\}&=E\{h(T)[E(T_1-h(T)\,|\,T)]\}\\&=0,\qquad なぜなら\ E(T_1\,|\,T=t)=h(t)\end{aligned} \tag{5a. 2. 8}$$

を得る．そうすると

$$\begin{aligned}E[T_1-g(\theta)]^2&=E[T_1-h(T)+h(T)-g(\theta)]^2\\&=E[T_1-h(T)]^2+E[h(T)-g(\theta)]^2\end{aligned}$$

$$\geqslant E[h(T)-g(\theta)]^2 \tag{5a.2.9}$$

を得る．なぜならば (5a.2.8) と (5a.2.7) を用いれば，

$$E[T_1-h(T)][h(T)-g(\theta)]=E\{[T_1-h(T)]h(T)\}-g(\theta)E[T_1-h(T)]$$
$$=0-0=0$$

だからである．(5a.2.9) の等号は，ほとんどいたるところで $T_1=h(T)$ であるときにだけ成立する．(5a.2.7) から，$E(T_1)=g(\theta)$ であれば $E[h(T)]=g(\theta)$ がわかる．

著者 Rao (1945d) および Blackwell (1947) によって証明された結果 (iii) は，十分統計量の重要な性質を示している．任意の統計量が与えられたとき，十分統計量の関数で，平均2乗誤差あるいは最小分散（偏りが許されない場合）の意味で，一様によい推定量をみいだすことができるのである．

注意 **1.**　十分統計量が完備であるとは，その十分統計量の関数で期待値が零であるのは，$\mathscr{P}_\theta$ に属する各々の測度についてほとんどいたるところ零という関数だけであることをいう．もしも完備十分統計量が存在すれば，そのあらゆる関数は，その期待値の一様 m.v.u.e. である．(iii) によれば，m.v. 推定量を求めるには，十分統計量の関数の集まりのなかでだけ捜せばよい．完備性の条件は，与えられた母数の関数にたいして，不偏である十分統計量の関数はただ1つであることを示す．このような場合，m.v.e. を求めるには，1つ不偏推定量を捜し，十分統計量を与えたときの，その推定量の条件つき期待値をとるだけで十分である (Rao(1946h, 1948d))．

注意 **2.**　(5a.2.7) で，T が十分統計量であることは，$h(T)$ が未知の母数と無関係な統計量であることを保証するためにだけ用いられていることがわかる．したがって，命題 (iii) は，T が十分統計量でなくとも，すべての t にたいして $E(T_1|T=t)$ が θ と無関係であるという性質をみたすような2つの統計量 (T_1,T) については成り立つ命題である．このことは Arnold and Katti (1972) によって指摘された．証明は同じである．

(iv)　平均2乗誤差の代わりに，任意の凸な損失関数 $W(T_1,\theta)$ を考えよう．イェンセンの不等式 (1e.5.6) を用いれば，$E[W(T_1,\theta)|T]\geqslant W(h(T),\theta)$ となる．それゆえ，T について期待値をとれば，

$$E[W(T_1,\theta)]\geqslant E[W(h(T),\theta)] \tag{5a.2.10}$$

を得る．これは，推定量と母数との差をより一般的な関数で測る場合への，(5a.2.9) の1つの拡張を与える．

例 **1.**　統計量 r を，成功の確率 π をもつ n 回のベルヌイ試行における成功の回数とする．$f(r)$ を，すべての π にたいして $E[f(r)]=0$ であるような関数とする．つまり

$$\sum f(r)\binom{n}{r}\pi^r(1-\pi)^{n-r}=0,\qquad 0<\pi<1 \text{ にたいして}$$

$$\Rightarrow \sum_0^n \binom{n}{r} f(r)x^r=0,\qquad \text{ここで } x=\frac{\pi}{1-\pi} \tag{5a.2.11}$$

である．x についての多項式 (5a.2.11) が，$0<x<\infty$ なるすべての x にたいして零ということから，その係数がすべて零であることがわかる．つまり，すべての r にたいして $f(r)=0$ である．したがって r は完備十分統計量であり，r の任意の関数は m.v.e. である．いま

$$E\left(\frac{r}{n}\right)=\pi,\qquad E\left(\frac{r}{n}\frac{r-1}{n-1}\right)=\pi^2,\qquad E\left(\frac{r}{n}\frac{n-r}{n-1}\right)=\pi(1-\pi)$$

であることは確かめられるので，期待値記号のなかの関数は，それぞれ対応する期待値の m.v. 推定量である．π だけではなく，π の推定量の分散の表現にでてくる $\pi(1-\pi)$ についても m.v.u. 推定ができるのである．ほかの例については，Lehmann and Scheffe (1950), Rao (1945d, 1946h, 1948d, 1949c) を参照せよ．

例 **2.**　Kolmogorov (1950) による (iii) の結果の大変興味深い応用は標本 $x_1,\cdots,x_n$ に基づいて，ある与えられた値 c を超える正規母集団の比率 π の不偏推定である．この場合には

$$\bar{x}=\frac{x_1+\cdots+x_n}{n}\ \text{と}\ S^2=(x_1-\bar{x})^2+\cdots+(x_n-\bar{x})^2$$

が，母集団の未知の平均 μ と分散 σ^2 の十分統計量である．これらが完備であることも確かめられる．次のような統計量を考える．

$$\left.\begin{aligned}T_1&=0, & x_1&\leqslant c\ \text{のとき}\\ &=1, & x_1&>c\ \text{のとき}\end{aligned}\right\}\qquad \text{ここで } x_1 \text{ は1番目の観測値である．}$$

これは，明らかに π の不偏推定量である．結果 (iii) を応用すれば，π の最小分散不偏推定量は $E(T_1|\bar{x},S)$ である．この統計量の具体的な値を求めるため，次のように考える．

$$E(T_1|\bar{x},S)=P\{x_1>c|\bar{x},S\}=P\left\{\frac{\sqrt{n}(x_1-\bar{x})}{\sqrt{n-1}S}>\frac{\sqrt{n}(c-\bar{x})}{\sqrt{n-1}S}\,\middle|\,\bar{x},S\right\}$$
$$=P\{u>u_0|\bar{x},S\},$$

ここで $u=\sqrt{n}(x_1-\bar{x})/\sqrt{n-1}S$ であり，$u_0=\sqrt{n}(c-\bar{x})/\sqrt{n-1}S$ である．しかしながら，確率変数 u は $\bar{x},S$ と独立に分布しており，3章の問題3で $m_1=1-n^{-1}$, $m_2=\cdots=m_n=-n^{-1}$ ととれば，その密度関数は const. $(1-u^2)^{(n-4)/2}$ で与えられる．それゆえ，

$$P[u>u_0|\bar{x},S]=\int_{u_0}^1 \text{const.}\,(1-u^2)^{(n-4)/2}du$$

となる．観測された u_0 の値に対応するこの積分の数値を求めるため，つまり，$\bar{x},S$ の観測値にたいして推定量 $E(T_1|\bar{x},S)$ の値を求めるために，ベータ分布 $B[1/2,(n-2)/2]$ の数表を用いてベータ変数が u_0^2 よりも大きい確率 P を決定する．u_0 が正であれば，求め

る推定値は $p/2$ であり，u_0 が負ならば，推定値は $1-p/2$ である．

（v） $\mathscr{P}_\theta$ は，σ 有限な測度 ν に関して，各 θ について確率密度（離散変数の場合には確率）$P(\cdot,\theta)$ をもつものとする．T を $g(\theta)$ の任意の不偏推定量とし，$A(\phi,\theta)=V[P(X,\phi)/P(X,\theta)|\theta]$，$S_\theta=\{x:P(x,\theta)>0\}$，$C=\{\phi:S_\phi\subset S_\theta\}$ とすると

$$V(T|\theta)\geqslant\sup_{\phi\in C}\frac{[g(\phi)-g(\theta)]^2}{A(\phi,\theta)} \tag{5a.2.12}$$

が得られる．

T は不偏推定量であるから，

$$\int_{S_\theta}TP(X,\phi)dv=g(\phi),\qquad \text{すべての}\,\phi\,\text{にたいして},$$

である．それゆえ

$$\int_{S_\theta}T\frac{P(X,\phi)-P(X,\theta)}{P(X,\theta)}P(X,\theta)dv=g(\phi)-g(\theta)$$

であり，したがって

$$\operatorname{cov}\left\{T,\frac{P(X,\phi)}{P(X,\theta)}-1\middle|\theta\right\}=g(\phi)-g(\theta)$$

となる．よって，C-S 不等式 $[\operatorname{cov}(x,y)]^2\leqslant V(x)V(y)$ を用いれば

$$V(T|\theta)\geqslant\frac{[g(\phi)-g(\theta)]^2}{V[P(X,\phi)/P(X,\theta)|\theta]},\qquad \text{すべての}\,\phi\in C\,\text{にたいして} \tag{5a.2.13}$$

が成り立つ．それゆえ (v) の (5a.2.12) は，何の正則条件もおかずに証明された．

以下では θ の近傍のすべての ϕ にたいして $S_\phi\subset S_\theta$ を仮定する．

（a） $g(\theta)$ が微分可能であるとし，その導関数を $g'(\theta)$ とする．また $\phi\to\theta$ のとき，$A(\phi,\theta)/(\phi-\theta)^2\to J(\theta)$ とする．そうすると，$V(T)\geqslant[g'(\theta)]^2/J(\theta)$ である．

この結果は (5a.2.13) の右辺の分母，分子を $(\phi-\theta)^2$ で割り，$\phi\to\theta$ のときの極限をとることにより得られる．

（b） $g(\theta)$ と $P(X,\theta)$ が微分可能であるとし，固定された θ と $|\phi-\theta|<\varepsilon$（与えられた正数とする）なる ϕ にたいして

$$\left|\frac{P(X,\phi)-P(X,\theta)}{(\phi-\theta)P(X,\theta)}\right|<G(X,\theta) \tag{5a.2.14}$$

であり，$E\{[G(X,\theta)]^2|\theta\}$ が存在するものとする．そうすると

$$V(T)\geqslant[g'(\theta)]^2/\mathscr{I}(\theta) \tag{5a.2.15}$$

である．ここで

$$\mathscr{I}(\theta)=\int\frac{1}{P(X,\theta)}\left(\frac{dP(X,\theta)}{d\theta}\right)^2dv$$

は，X の標本空間からの1つの観測値に含まれる，母数 θ についての，フィッシャーの情報の尺度である［$\mathscr{I}(\theta)$ についてのくわしい議論は，5a.4 を参照せよ］．

仮定された条件のもとでは

$$J(\theta)=\lim_{\phi\to\theta}\frac{A(\phi,\theta)}{(\phi-\theta)^2}=\lim_{\phi\to\theta}\int\frac{[P(X,\phi)-P(X,\theta)]^2}{[(\phi-\theta)P(X,\theta)]^2}P(X,\theta)dv$$
$$=\int\left(\frac{P'(X,\theta)}{P(X,\theta)}\right)^2P(X,\theta)dv=\mathscr{I}(\theta) \tag{5a.2.16}$$

となる．ここで，(5a.2.16) の極限は，(5a.2.14) を用いてルベーグの上限収束定理により，積分記号のなかでとっている．(5a.2.14) の条件のほかにも，極限を積分記号のなかでとるための十分条件はあるだろう．

それゆえ，$V(T)\geqslant[g'(\theta)]^2/J(\theta)$ の $J(\theta)$ に $\mathscr{I}(\theta)$ を代入すれば，結果 (5a.2.15) が得られる．(5a.2.15) は情報の限界と呼んでよいであろう (Cramer (1946), Rao (1945d) を参照せよ)．(5a.2.12) の形での不等式は Chapman and Robbins (1951) による．ほかの形の不等式もいくつか与えられている (Bhattacharya (1946), Gart (1959), Hodges and Lehmann (1951), Stein (1950) を参照せよ)．

(5a.2.15) は (5a.2.14) を用いて，直接的にも証明できる．T は $g(\theta)$ の不偏推定量であるから，

$$\int T\frac{P(X,\phi)-P(X,\theta)}{(\phi-\theta)P(X,\theta)}P(X,\theta)dv=\frac{g(\phi)-g(\theta)}{\phi-\theta} \tag{5a.2.17}$$

となる．$V(T)$ が有限であるとする．いま

$$\left|T\frac{P(X,\phi)-P(X,\theta)}{(\phi-\theta)P(X,\theta)}\right|<|TG(X,\theta)|$$

であり，$E(T^2)$ と $E[G(X,\theta)]^2$ は存在するので $E[|TG(X,\theta)|]$ が存在する．したがって，(5a.2.17) の両辺の $\phi\to\theta$ のときの極限をとれば，

$$\int T\frac{P'(X,\theta)}{P(X,\theta)}P(X,\theta)dv=g'(\theta)$$

である．つまり，$\operatorname{cov}[T,P'(X,\theta)/P(X,\theta)]=g'(\theta)$ であることがわかる．それゆえ，$V(T)\geqslant[g'(\theta)]^2/\mathscr{I}(\theta)$ を得る．

(5a.2.15) で等号が成り立つときは，T と $P'(\theta)/P(\theta)$ の相関係数は1である．その場合，

$$T=\lambda\frac{P'}{P}+g(\theta),\qquad (\lambda\,\text{は}\,X\,\text{に無関係}) \tag{5a.2.18}$$

が，$P(\cdot,\theta)$ に関して確率測度零の X の集合を除いて，成立する．(5a.2.18) を θ について積分すれば，$P(X,\theta)$ は

$$\exp[g_1(\theta)T(X)+g_2(\theta)+U(X)]$$

という形であることがわかり，これは T が θ の十分統計量であることを示している．

例 1. 確率密度 $p(x,\theta)$ をもつ母集団からの n 個の独立な観測値 $x_1,\cdots,x_n$ を考え

よう．このような場合 $P(X,\theta)=p(x_1,\theta)\cdots p(x_n,\theta)$ となり，$\mathscr{I}(\theta)=ni(\theta)$ となる．ここで

$$i(\theta)=\int\left[\frac{p'(x,\theta)}{p(x,\theta)}\right]^2 p(x,\theta)dx$$

である．$p(x,\theta)=N(x|\theta,\sigma^2)$ のときには，$i(\theta)=1/\sigma^2$ となることが容易に確かめられるので，θ の不偏推定量の分散の下界は $1/ni=\sigma^2/n$ である．ところで，統計量 $\bar{x}$ は θ の不偏推定量であり，

$$V(\bar{x})=\frac{\sigma^2}{n}=\frac{1}{\mathscr{I}(\theta)}$$

であるから，$\bar{x}$ は θ の m.v.u.e. である．

例 2.　$p(x,\theta)$ がガンマ分布の密度関数

$$\frac{\theta^k}{\Gamma(k)}e^{-\theta x}x^{k-1}$$

であるときは，$i(\theta)=k/\theta^2$ である．統計量 $(\bar{x}/k)$ は $1/\theta$ の不偏推定量であり，

$$V\left(\frac{\bar{x}}{k}\right)=\frac{1}{kn\theta^2}=\left[\frac{d(1/\theta)}{d\theta}\right]^2\frac{1}{\mathscr{I}(\theta)}$$

であるから，$\bar{x}/k$ は $1/\theta$ の m.v.u.e. である．

5a.3　母数が複数個ある場合

$P(X,\theta)$ の θ がベクトルであるとし，

$$P_{ij}=-\frac{\partial^2\log P}{\partial\theta_i\partial\theta_j}\quad かつ\quad E(P_{ij})=\mathscr{I}_{ij}$$

とする．行列 $\mathscr{I}=(\mathscr{I}_{ij})$, $i,j=1,\cdots,k$ は情報行列と呼ばれる．T を θ の十分統計量とする．さて，次の主要な結果を証明しよう．

(i)　r 個の統計量 $f_1,\cdots,f_r$ は

$$\left.\begin{array}{ll}(\mathrm{a}) & E(f_i)=g_i(\theta_1,\cdots,\theta_k)\\ (\mathrm{b}) & E[(f_i-g_i)(f_j-g_j)]=V_{ij}\end{array}\right\}\quad i,j=1,\cdots,r$$

であるものとする．そうすると次のことが成り立つ．

(A)　次のような十分統計量 T の関数 $m_1,\cdots,m_r$ が存在する．

(A.1)　$E(m_i)=g_i(\theta_1,\cdots,\theta_k)$,

(A.2)　$U_{ij}=E[(m_i-g_i)(m_j-g_j)]$ とし，$\mathbf{U}=(U_{ij})$, $\mathbf{V}=(V_{ij})$ とすれば，行列 $\mathbf{V}-\mathbf{U}$ が非負定符号である．

(B)　いま

$$\frac{\partial}{\partial\theta_j}\int f_iP(X,\theta)dv=\int f_i\frac{\partial P(X,\theta)}{\partial\theta_j}dv=\frac{\partial g_i}{\partial\theta_j}$$

とし，Δ を行列 $(\partial g_i/\partial\theta_j)$, $i=1,\cdots,r$, $j=1,\cdots,k$ とする．そうすると $\mathbf{V}-\Delta\mathscr{I}^{-1}\Delta'$ は非負定符号である．ここで $\mathscr{I}^{-1}$ は情報行列 $\mathscr{I}$ の逆行列である [$\mathscr{I}$ が特異であるならば $\mathscr{I}^{-1}$ は g 逆行列 $\mathscr{I}^-$ におきかえられる]．

(A) を証明するために線形関数 $\mathbf{L'F}$ を考える．ここで $\mathbf{F}'=(f_1,\cdots,f_r)$ であり $\mathbf{L}$ は1つの定数ベクトルである．さらに $E(f_i|T)=m_i$ とする．このとき，$\mathbf{M}'=(m_1,\cdots,m_r)$ とすれば，$E(\mathbf{L'F}|T)=\mathbf{L'M}$ である．結果 [(iii)，5a.2] を関数 $\mathbf{L'F}$ に適用すれば，$V(\mathbf{L'F})\geqslant V(\mathbf{L'M})$ であるから，

$$\mathbf{L'VL}\geqslant\mathbf{L'UL}\quad つまり\quad \mathbf{L}'(\mathbf{V}-\mathbf{U})\mathbf{L}\geqslant 0$$

が任意の定数ベクトル $\mathbf{L}$ について成り立つ．したがって，$\mathbf{V}-\mathbf{U}$ は非負定符号であり，(A) が証明された．

(B) を証明するために，関係式

$$\int f_i\frac{\partial P}{\partial\theta_j}dv=\frac{\partial g_i}{\partial\theta_j}$$

に注意し，$f_1,\cdots,f_r$, $(1/P)(\partial P/\partial\theta_1),\cdots,(1/P)(\partial P/\partial\theta_k)$ の分散共分散行列を考えれば，次のような分割行列の形に書くことができる．

$$\begin{pmatrix}\mathbf{V} & \Delta\\ \Delta' & \mathscr{I}\end{pmatrix}$$

行列式（いま単位行列を $\mathbf{I}$ であらわす）

$$\begin{vmatrix}\mathbf{I} & -\Delta\mathscr{I}^{-1}\\ 0 & \mathscr{I}^{-1}\end{vmatrix}\quad と\quad \begin{vmatrix}\mathbf{V} & \Delta\\ \Delta' & \mathscr{I}\end{vmatrix}$$

は非負であるから，それらの積

$$\begin{vmatrix}\mathbf{V}-\Delta\mathscr{I}^{-1}\Delta' & 0\\ \mathscr{I}^{-1}\Delta' & \mathbf{I}\end{vmatrix}=|\mathbf{V}-\Delta\mathscr{I}^{-1}\Delta'|$$

も非負である．これは，統計量 $f_1,\cdots,f_r$ の任意の部分集合についても成り立つので，$\mathbf{V}-\Delta\mathscr{I}^{-1}\Delta'$ は非負定符号である．これで (B) が証明された．

主要な結果 (A) と (B) から一連の結果が得られる．

(a)　$\mathbf{V}-\Delta\mathscr{I}^{-1}\Delta'$ の対角要素だけを考えれば

$$V_{ii}\geqslant\sum\sum\mathscr{I}^{mn}\frac{\partial g_i}{\partial\theta_m}\frac{\partial g_i}{\partial\theta_n}\tag{5a.3.1}$$

である．ここで $\mathscr{I}^{mn}$ は情報行列 $(\mathscr{I}_{mn})$ の逆行列の要素である．

この結果は g_i の推定量の分散が，推定の方法とは無関係に定義された量よりも小さくはないことを示している．これは (5a.2.15) で導かれた表現の，母数が複数個ある場合への一般化である．$g_i=\theta_i$ $(i=1,\cdots,k)$ のときには，関係式 (5a.3.1) は

$$V_{ii}\geqslant\mathscr{I}^{ii}\tag{5a.3.2}$$

となる．θ_i の推定値について (5a. 2. 15) から得られる下界は $1/\mathscr{I}_{ii}$ であるが，$\mathscr{I}^{ii}$ はこれより小さくないことに注意せよ．$\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_k$ の値がわかっていれば，(5a. 2. 15) で与えられる限界が適用できる．そうでなければ，θ_i の推定値はほかの母数と無関係でなければならず，そのために限界は大きくなるだろう．

（b） 行列 $\mathbf{V}-\mathbf{U}$ は非負定符号であるから，$V_{ii} \geqslant U_{ii}$ $(i=1, \cdots, r)$ である．このことは，到達可能な最小の分散をもつ推定量は十分統計量の関数であることを示している．

（c） 行列 $\mathbf{V}-\Delta\mathscr{I}^{-1}\Delta'$ は非負定符号であるから，

$$|\mathbf{V}| \geqslant |\Delta\mathscr{I}^{-1}\Delta'| \tag{5a. 3. 3}$$

である．

$|\mathbf{V}|$ という量は，推定量の一般化分散と呼ばれている．結果 (5a. 3. 3) は，この一般化分散が，推定の方法とは無関係に定義された量よりも小さくはないことを示している．

（d） $\mathbf{V}-\mathbf{U}$ は非負定符号であるから，$|\mathbf{V}| \geqslant |\mathbf{U}|$ である．

これは，最小の一般化分散をもつ推定量も十分統計量の関数であることを示している．

（e） もしも

$$V_{ii}=\sum\sum\mathscr{I}^{mn}\frac{\partial g_i}{\partial\theta_m}\frac{\partial g_i}{\partial\theta_n} \tag{5a. 3. 4}$$

であれば，g_i の推定量は最小分散をもつが，そのとき

$$V_{ij}=\sum\sum\mathscr{I}^{mn}\frac{\partial g_i}{\partial\theta_m}\frac{\partial g_j}{\partial\theta_n} \tag{5a. 3. 5}$$

となる．つまり m. v. u. e. と，他の任意の母数の関数の，任意の不偏推定量との共分散は，推定の方法とは無関係に定義された一定値である．

この結果は，$|\mathbf{V}-\Delta\mathscr{I}^{-1}\Delta'|$ の部分行列式である行列式

$$\begin{vmatrix} V_{ii}-\sum\sum\mathscr{I}^{mn}\frac{\partial g_i}{\partial\theta_m}\frac{\partial g_i}{\partial\theta_n} & V_{ij}-\sum\sum\mathscr{I}^{mn}\frac{\partial g_i}{\partial\theta_m}\frac{\partial g_j}{\partial\theta_n} \\ V_{ij}-\sum\sum\mathscr{I}^{mn}\frac{\partial g_i}{\partial\theta_m}\frac{\partial g_j}{\partial\theta_n} & V_{jj}-\sum\sum\mathscr{I}^{mn}\frac{\partial g_j}{\partial\theta_m}\frac{\partial g_j}{\partial\theta_n} \end{vmatrix}$$

が非負であることから導かれる．

例　正規母集団 $N(\theta, \sigma^2)$ からの n 個の独立な観測値を考える．θ と σ^2 の情報行列は

$$\mathscr{I}=\begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{pmatrix}$$

であることが容易に確かめられる．したがって，その逆行列は

$$\mathscr{I}^{-1}=\begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}$$

である．θ の推定量としての $\bar{x}$ は到達可能な最小の分散をもち，$\mathscr{I}^{\theta\sigma^2}=0$ であるから，

σ^2 の任意の不偏推定量は $\bar{x}$ と無相関である．この結果は多変量正規分布の場合に拡張ができ，標本平均は分散，共分散のすべての不偏推定量と無相関であることが示される．

σ^2 の推定のために，統計量 $s^2=\sum(x_i-\bar{x})^2/(n-1)$ を考えれば，

$$E(s^2)=\sigma^2, \qquad V(s^2)=\frac{2\sigma^4}{n-1} \tag{5a. 3. 6}$$

であるので，分散の下界 $2\sigma^4/n$ に到達していない．しかし実は，下に示すように，これが到達できる最小の値である．σ^2 の任意の不偏推定量を考え，s^2 との差を $f(s, \bar{x})$ とすれば，

$$\int f(s, \bar{x})\exp\{-[n(\bar{x}-\theta)^2+(n-1)s^2]/2\sigma^2\}\,dv=0$$

である．θ について2回微分すれば，

$$E[(\bar{x}-\theta)^2 f(s, \bar{x})]=0$$

が導かれる．一方 σ^2 について微分すれば，

$$E[n(\bar{x}-\theta)^2 f(s, \bar{x})+(n-1)s^2 f(s, \bar{x})]=0$$

を得る．上の2つを合わせれば $\mathrm{cov}[s^2, f(s, \bar{x})]=0$ であるので，

$$V[s^2+f(\bar{x}, s)]=V(s^2)+V[f(s, \bar{x})]$$

である．つまり $V(s^2)$ が，σ^2 の不偏推定量の到達できる最小の分散である．それゆえ，$\bar{x}$ と s^2 が θ と σ^2 の m. v. 不偏推定量である．

5a. 4　フィッシャーの情報の尺度

X をベクトル値の，あるいはより一般の確率変数とし，値域を S とし，σ 有限な測度 v について確率密度 $P(\cdot, \theta)$ をもつものとする．さらに $P(\cdot, \theta)$ は θ について微分可能とし，任意の可測集合 $C\subset S$ にたいして

$$\frac{d}{d\theta}\int_C P(X, \theta)\,dv=\int_C\frac{dP(X, \theta)}{d\theta}\,dv \tag{5a. 4. 1}$$

であるものとする．確率変数 X に含まれる θ に関するフィッシャーの情報の尺度は次のように定義される．

$$\mathscr{I}(\theta)=E\left(\frac{d\log P}{d\theta}\right)^2=V\left(\frac{d\log P}{d\theta}\right), \quad \text{なぜなら}\ E\left(\frac{d\log P}{d\theta}\right)=0\ \text{だから}. \tag{5a. 4. 2}$$

情報の尺度 (5a. 4. 2) のいくつかの性質を調べる．

（i） $\mathscr{I}_1, \mathscr{I}_2$ を，それぞれ2つの独立な確率変数 X_1, X_2 に含まれる情報とし，$\mathscr{I}$ を (X_1, X_2) に含まれる情報とする．そうすると $\mathscr{I}=\mathscr{I}_1+\mathscr{I}_2$ である．

$P_1(\cdot, \theta)$ と $P_2(\cdot, \theta)$ をそれぞれ X_1, X_2 の密度とすれば，(X_1, X_2) の同時密度は，$P_1(\cdot, \theta)P_2(\cdot, \theta)$ である．定義より

$$\mathscr{I} = E\left\{\frac{d}{d\theta}[\log P_1(X_1,\theta)P_2(X_2,\theta)]\right\}^2$$
$$= E\left(\frac{d\log P_1}{d\theta}\right)^2 + E\left(\frac{d\log P_2}{d\theta}\right)^2 + 2E\left(\frac{d\log P_1}{d\theta}\cdot\frac{d\log P_2}{d\theta}\right)$$
$$= \mathscr{I}_1 + \mathscr{I}_2$$

である．なぜならば

$$E\left(\frac{d\log P_1}{d\theta}\cdot\frac{d\log P_2}{d\theta}\right) = E\left(\frac{d\log P_1}{d\theta}\right)E\left(\frac{d\log P_2}{d\theta}\right) = 0$$

だからである．

（ii）$X_1,\cdots,X_k$ を，独立で同一分布に従う確率変数とし，$\mathscr{I}$ を各々に含まれる情報とする．そうすると $(X_1,\cdots,X_k)$ に含まれる情報は $k\mathscr{I}$ である．

（iii）T を X の可測関数とし，σ 有限な測度 ν に関して，密度関数 $\phi(\cdot,\theta)$ をもち，(5a. 4. 1) と同様の条件をみたすものとする．さらに，$\mathscr{I}_T$ を T に含まれる情報，つまり

$$\mathscr{I}_T = E[\phi'(T,\theta)/\phi(T,\theta)]^2$$

とする．そうすると

（a）
$$E\left(\frac{P'(X,\theta)}{P(X,\theta)}\middle| T=t\right) = \frac{\phi'(t,\theta)}{\phi(t,\theta)},$$

（b）
$$\mathscr{I} \geqq \mathscr{I}_T \qquad (5a.4.3)$$

である．

Aを T の空間の可測集合とし，A' をそれに対応する X の空間の集合とする．そうすると（便宜上密度関数のなかで同じ記号 X と T を用いれば），

$$\int_{A'}\frac{P'(X,\theta)}{P(X,\theta)}\cdot P(X,\theta)d\nu = \frac{d}{d\theta}\int_{A'}P(X,\theta)d\nu$$
$$= \frac{d}{d\theta}\int_A \phi(T,\theta)d\nu = \int_A \phi'(T,\theta)d\nu = \int_A \frac{\phi'(T,\theta)}{\phi(T,\theta)}\cdot\phi(T,\theta)d\nu$$

が任意のAにたいして成り立つ．したがって条件つき期待値の定義から，

$$E\left(\frac{P'(X,\theta)}{P(X,\theta)}\middle| T=t\right) = \frac{\phi'(t,\theta)}{\phi(t,\theta)}$$

が導かれる．いま

$$E\left\{\frac{\phi'(T,\theta)}{\phi(T,\theta)}\cdot\frac{P'(X,\theta)}{P(X,\theta)}\right\} = E_T\left\{\frac{\phi'(T,\theta)}{\phi(T,\theta)}E\left[\frac{P'(X,\theta)}{P(X,\theta)}\middle| T\right]\right\} \quad \text{(2b. 3. 7) を用いた}$$
$$= E_T\left[\frac{\phi'(T,\theta)}{\phi(T,\theta)}\right]^2 = \mathscr{I}_T$$

である．したがって

$$E\left(\frac{P'}{P}-\frac{\phi'}{\phi}\right)^2 = E\left(\frac{P'}{P}\right)^2 + E\left(\frac{\phi'}{\phi}\right)^2 - 2E\left(\frac{P'}{P}\cdot\frac{\phi'}{\phi}\right)$$
$$= \mathscr{I} + \mathscr{I}_T - 2\mathscr{I}_T = \mathscr{I} - \mathscr{I}_T \geqq 0$$

である．

X が全標本をあらわし，T が1つの統計量をあらわすとすると，結果（iii, b）は，一般には X を T でおきかえることにより情報の損失があることを示している．情報の損失がないとき，つまり，$\mathscr{I}=\mathscr{I}_T$ のときには，

$$E\left(\frac{P'}{P}-\frac{\phi'}{\phi}\right)^2 = 0 \Rightarrow \frac{P'}{P} = \frac{\phi'}{\phi} \qquad \text{a. e. } (\nu),$$

つまり，$P(X,\theta)=\phi(T,\theta)\Psi(X,T)$ となる．ここで Ψ は θ に無関係である．したがって T は θ の十分統計量である．

（iv）母数が複数個の場合には，

$$\mathscr{I}_{rs} = E\left(\frac{\partial\log P}{\partial\theta_r}\cdot\frac{\partial\log P}{\partial\theta_s}\right), \qquad r,s=1,\cdots,k$$

とする．そのとき，行列 $\mathscr{I}=(\mathscr{I}_{rs})$ を情報行列と定義する．

1母数の場合と同様にして次のことが容易に証明できる．

（a）$\mathscr{I}=\mathscr{I}_1+\mathscr{I}_2$ である．ここで $\mathscr{I}_1,\mathscr{I}_2$ はそれぞれ2つの独立な確率変数 X_1,X_2 からつくられる情報行列，$\mathscr{I}$ は (X_1,X_2) からつくられる情報行列である．

（b）行列 $\mathscr{I}-\mathscr{I}_T$ は非負定符号（正あるいは半正定符号）である．ここで $\mathscr{I}_T$ は X の関数 T からつくられる情報行列である．

分布の内在的精度の尺度としての情報　確率変数あるいはその分布に含まれる，未知母数 θ についての情報というとき，その確率変数の観測値を得ることにより，どれくらい未知の値 θ についての不確定性が減少するかの度合いを意味する．母数の各々の値に対応して確率1でただ1つの観測値が存在するならば，確率変数は最大の情報をもっているのである．一方，確率変数が，母数のすべての値にたいして同じ分布をもつならば，このような確率変数の観測値をもとにしては，θ について何も述べることができない．確率変数の母数に関する感度は，母数の値が変化したときに分布がどの程度変わるかによって判断されるだろう．母数の2つの値 θ と θ' についての，X の確率密度を $P(\cdot,\theta)$ と $P(\cdot,\theta')$ であらわせば，2つの分布の差は，ヘリンガーの距離

$$\cos^{-1}\int\sqrt{P(X,\theta)P(X,\theta')}d\nu \qquad (5a.4.4)$$

のような，ある距離関数によって測ることができるだろう．$\theta'=\theta+\delta\theta$ とし，$P(X,\theta')$ をテイラー展開し，$\delta\theta$ の高次の項を無視すれば，(5a. 4. 4) は

$$\cos^{-1}\int P(\theta)\left\{1-\frac{1}{8}\left[\frac{P'(\theta)}{P(\theta)}\right]^2(\delta\theta)^2\right\}d\nu = \cos^{-1}\left[1-\frac{1}{8}\mathscr{I}(\theta)(\delta\theta)^2\right] \qquad (5a.4.5)$$

となる．ここで $\mathscr{I}(\theta)$ はフィッシャーの情報である．$\mathscr{I}(\theta)$ は正であるから，距離 (5a. 4. 5) は，$\mathscr{I}(\theta)$ の値が増加すれば増加する．したがって $\mathscr{I}(\theta)$ は，母数の値の微小変化に関する確率変数の感度の尺度になっている．

読者は，さまざまな距離の尺度について，フィッシャーの情報が，母数の値の微小変化にたいする感度の指標になっていることを検証できるだろう（Rao (1962e, g) 参照）．母数が複数個ある場合には，母数 $(\theta_1, \theta_2, \cdots)$ と $(\theta_1+\delta\theta_1, \theta_2+\delta\theta_2, \cdots)$ にたいする確率変数の分布の間の距離として，(5a. 4. 5) の $\mathscr{I}(\theta)(\delta\theta)^2$ の代わりに2次微分形式

$$\sum\sum \mathscr{I}_{ij}\delta\theta_i\delta\theta_j$$

を得る．そのとき，母数の変化に関する確率変数の感度は，情報行列を全体として調べることにより判断できるだろう (Rao (1945d) 参照).

5a. 5　不偏推定量の改良

4k 節で，線形模型において，母数空間の先験的に与えられたある値の周辺の領域では，BLUE よりも平均2乗誤差が小さいような偏りのある推定量が存在しうることを示した．実際のところ，未知母数についてのなんらかの事前情報があれば，不偏推定量を改良することは常に可能なことである．

正規分布の未知の平均 μ を推定するものとすると，標本平均 $\bar{X}$ が十分統計量であり，最小分散不偏推定量である．もしも $\mu\in(a, b)$ ということがわかっていれば，別の推定量として

$$T=\begin{cases} a & \bar{X}<a \text{ のとき} \\ \bar{X} & a\leqslant\bar{X}\leqslant b \text{ のとき} \\ b & \bar{X}>b \text{ のとき}\end{cases} \tag{5a. 5. 1}$$

を考えることができる．T が μ の不偏推定量ではないことは容易にわかるが，

$$E(T-\mu)^2<E(\bar{X}-\mu)^2, \qquad a<\mu<b \text{ のとき},$$

であるので，$\mu\in(a, b)$ という事前情報が正しければ，T は μ の推定量として $\bar{X}$ よりも一様によい推定量である．

$\mathbf{T}'=(T_1, \cdots, T_k)$ は $\boldsymbol{\tau}_\theta'=(\tau_1(\theta), \cdots, \tau_k(\theta))$ の不偏推定量であり，正則な共分散行列 $\mathbf{V}_\theta$ をもつものとする．場合によっては

$$\mathbf{V}_\theta-E_\theta[(b\mathbf{T}-\boldsymbol{\tau}_\theta)(b\mathbf{T}-\boldsymbol{\tau}_\theta)'] \tag{5a. 5. 2}$$

が，未知母数 θ の全範囲について非負定符号になるように，θ と無関係な定数 b を定めることが可能である．ここで (5a. 5. 2) の第2項は $b\mathbf{T}$ の $\boldsymbol{\tau}_\theta$ からの偏差の平均である．もしもそうであれば，$b\mathbf{T}$ は任意の2次損失関数のもとで $\mathbf{T}$ よりも一様によい $\boldsymbol{\tau}_\theta$ の推定量である．このような改良ができる例はいくつか知られており，Perlman (1972) により示された次の命題が一般的結果を与えている．

(5a. 5. 2) が非負定符号になるような定数 b $(0<b<1)$ が存在するための必要十分条件は，$\boldsymbol{\tau}_\theta'\mathbf{V}_\theta^{-1}\boldsymbol{\tau}_\theta$ が θ の全範囲にわたって有界であることである．

(5a. 5. 2) を $\mathbf{V}_\theta-b^2\mathbf{V}_\theta-(1-b)^2\boldsymbol{\tau}_\theta\boldsymbol{\tau}_\theta'$ と書きかえれば，これが非負定符号であるための必要十分条件は，$m\mathbf{V}_\theta-\boldsymbol{\tau}_\theta\boldsymbol{\tau}_\theta'$ が非負定符号であることである．ここで $m=(1+b)/(1-b)$ である．$m\mathbf{V}_\theta-\boldsymbol{\tau}_\theta\boldsymbol{\tau}_\theta'$ が非負定符号であれば，C-S 等式 (51ページ) を用いることにより，

$$m\geqslant\sup_{\mathbf{x}}\frac{\mathbf{x}'\boldsymbol{\tau}_\theta\boldsymbol{\tau}_\theta'\mathbf{x}}{\mathbf{x}'\mathbf{V}_\theta\mathbf{x}}=\boldsymbol{\tau}_\theta'\mathbf{V}_\theta^{-1}\boldsymbol{\tau}_\theta \tag{5a. 5. 3}$$

となる．したがって，定数 b, $0<b<1$ が存在すれば $\boldsymbol{\tau}_\theta'\mathbf{V}_\theta^{-1}\boldsymbol{\tau}_\theta$ はすべての θ にたいして有界でなければならない．逆に，$\sup\boldsymbol{\tau}_\theta'\mathbf{V}_\theta^{-1}\boldsymbol{\tau}_\theta=m_0<\infty$ であれば，$0<b<1$ でありかつ $b>(m_0-1)/(m_0+1)$ である任意の b にたいして，(5a. 5. 2) は非負定符号となる．

(5a. 5. 2) で与えられる行列は零行列にはなりえないことに注意せよ．なぜなら，もし零行列であれば，$m\mathbf{I}-\mathbf{V}_\theta^{-1/2}\boldsymbol{\tau}_\theta\boldsymbol{\tau}_\theta'\mathbf{V}_\theta^{-1/2}$ も零行列になるが，これは $\mathbf{V}_\theta^{-1/2}\boldsymbol{\tau}_\theta\boldsymbol{\tau}_\theta'\mathbf{V}_\theta^{-1/2}$ の階数が1であるので起こりえないからである．

このように，事前情報 (5a. 5. 3) があれば，不偏推定量を一様に改良することができる．

読者は次のことを検証することができるだろう．$E(y)=\mu$, $V(y)=\sigma^2$, $\theta=\sigma/\mu$ のとき，もしも $c^2(1-\theta^2)\leqslant 2\theta^2$ であれば，すべての μ にたいして

$$E(y-\mu)^2\geqslant E[(1+c^2)^{-1}y-\mu]^2 \tag{5a. 5. 4}$$

が成り立つ．変動係数 θ のおおよその値がわかっていれば，c を適当に選んで不偏推定量 y を改良することが可能になる．c を θ に等しくとれれば最も好都合である．

5b　一般的手順

5b. 1　一般的問題の記述（ベイズの定理）

確率空間 $(\Omega, \mathscr{B})$ と確率変数（ω の関数）

$$\theta_1, \cdots, \theta_m\,;\, X_1, \cdots, X_n$$

を考える．ここで θ_i は仮説的（観測不可能）であり X_j は観測可能である．実際的な興味のある推測問題は，観測不可能な変量を，観測可能な変量によって予測することである．この一般的問題では，古典的な回帰問題の場合のように，変量 X_j のベクトル $\mathbf{X}$ が与えられたときの，変量 θ_i のベクトル $\boldsymbol{\theta}$ の条件つき分布を研究することが必要であるのは明らかである．$(\boldsymbol{\theta}, \mathbf{X})$ の同時分布がどの程度指定されるかによって，さまざまな状況を考えられる．

問題を単純にし，条件つき分布に関するある種のむずかしさを避けるため，$\boldsymbol{\theta}$ が与えられたときの $\mathbf{X}$ の条件つき確率密度が存在するものとし，それを $P(\cdot|\cdot)$ であらわすことにする．F が $\boldsymbol{\theta}$ の周辺分布関数であるとすれば，$\mathbf{X}$ の $\mathbf{X}_0$ での周辺確率密度は

$$G(\mathbf{X}_0) = \int P(\mathbf{X}_0|\theta)dF(\theta) \tag{5b.1.1}$$

であり，$\mathbf{X} = \mathbf{X}_0$ が与えられたときの θ の条件つき確率分布は

$$dF(\cdot|\mathbf{X}_0) = \frac{P(\mathbf{X}_0|\cdot)}{G(\mathbf{X}_0)}dF(\cdot) \tag{5b.1.2}$$

である．

分布関数 F は θ の事前分布関数と呼ばれ，$F(\cdot|\mathbf{X}_0)$ は $\mathbf{X} = \mathbf{X}_0$ が与えられたときの θ の事後分布関数と呼ばれる．これらの用語は，上に述べたような一般的問題では，特別な意味があるわけではない．いくつか例を考えよう．

結果 (5b.1.2) はベイズの定理として知られている．

5b.2　$(\theta, \mathbf{X})$ の同時分布関数が完全に知られている場合

$(\theta, \mathbf{X})$ の同時分布が完全に知られているときには，$\mathbf{X}$ を観測したときの θ の不確定性を表現するのに，表現 (5b.1.2) が申し分のない形である．$\mathbf{X} = \mathbf{X}_0$ が与えられたとき，θ が与えられた領域 A にはいる確率を定めるには，

$$\int_A dF(\theta|\mathbf{X}_0) = \int_A \frac{P(\mathbf{X}_0|\theta)}{G(\mathbf{X}_0)}dF(\theta) \tag{5b.2.1}$$

の値を計算すればよい．θ の関数 $g(\theta)$ について $1-\alpha$ 信頼区間を求めたいのならば，

$$\int_{a<g(\theta)<b} \frac{P(\mathbf{X}_0|\theta)}{G(\mathbf{X}_0)}dF(\theta) = 1-\alpha \tag{5b.2.2}$$

という条件のもとで，$b-a$ が最小になるように (a, b) を決定すればよい．$g(\theta)$ の点推定値は，条件つき分布に関するその期待値

$$\int g(\theta)dF(\theta|\mathbf{X}_0) = \int g(\theta)\frac{P(\mathbf{X}_0|\theta)}{G(\mathbf{X}_0)}dF(\theta) \tag{5b.2.3}$$

によって与えられる．これは，$g(\theta)$ の $\mathbf{X}$ への（$\mathbf{X}_0$ での）回帰であり，平均2乗誤差を最小にする（(4g.1.1) を参照せよ)．

例として，n 回のベルヌイ試行で r 回の成功が観測されたということをもとにして，1回の試行で成功する確率 π を予測することを考えよう．π の事前（周辺）確率密度を

$$\frac{\pi^{\alpha-1}(1-\pi)^{\gamma-1}}{\beta(\alpha,\gamma)}, \quad 0<\pi<1 \tag{5b.2.4}$$

とすれば，r と π の同時確率分布は

$$\binom{n}{r}\pi^r(1-\pi)^{n-r}\frac{\pi^{\alpha-1}(1-\pi)^{\gamma-1}}{\beta(\alpha,\gamma)}d\pi \tag{5b.2.5}$$

で与えられ，r だけの分布は，π について積分することにより，

$$\binom{n}{r}\frac{\beta(\alpha+r, \gamma+n-r)}{\beta(\alpha,\gamma)} \tag{5b.2.6}$$

である．ここで $\beta(p, q)$ はベータ関数である．r が与えられたときの π の事後分布は (5b.2.5)/(5b.2.6) で与えられるから，

$$\frac{\pi^{\alpha+r-1}(1-\pi)^{\gamma+n-r-1}}{\beta(\alpha+r, \gamma+n-r)}d\pi \tag{5b.2.7}$$

である．推定値として π のどんな値を選んでも，その結果は，分布 (5b.2.7) を用いて完全に調べることができる．

π の点推定値が欲しいならば，分布 (5b.2.7) の平均あるいはモードをとればよいだろう．平均は，与えられた r にたいして平均2乗誤差を最小にするが，この場合

$$\frac{\alpha+r}{\alpha+\gamma+n} \tag{5b.2.8}$$

で与えられる．これは平均2乗誤差最小の規準と事前分布 (5b.2.4) に関するベイズ推定量と呼ぶことができる．推定量 (5b.2.8) は，伝統的な推定量 r/n とは異なることがわかる．

さらに m 回試行を繰り返したときに s 回成功する確率を決定するのにも，分布 (5b.2.7) を用いることができる．与えられた π については，s 回成功する確率は

$$[m!/s!(m-s)!]\pi^s(1-\pi)^{m-s}$$

である．これを事後分布 (5b.2.7) について積分すれば，r が与えられたときの s の確率分布

$$\frac{m!}{(m-s)!s!}\frac{\beta(\alpha+r+s, \gamma+m+n-r-s)}{\beta(\alpha+r, \gamma+n-r)} \tag{5b.2.9}$$

が得られる．

5b.3　同程度に不明とみなす法則

残念ながら，実際の場面では $(\theta, \mathbf{X})$ の同時分布関数が完全にわかることはまれである．θ を与えたときの $\mathbf{X}$ の条件つき密度 $P(\cdot|\cdot)$ の関数形は，少なくともかなりのよさの近似をもって指定することができるが，事前分布 F を与えることは困難な問題である．θ の事前分布について何の知識もない場合には，事前分布を選択する何らかの一貫した規則が与えられるのではないかと提案がされている．Jeffreys (1939) 等によってこのような方向への試みがなされているが，有用な結果は得られていないようである．

5b.2 で議論した2項分布の母数の推定問題に，同程度に不明とみなす法則を用いてみよう．すると π に一様分布を仮定することになり，これは (5b.2.4) で $\alpha=1$, $\gamma=1$ とした特別な場合になる．この場合，r 回成功したということが与えられたときの π の事後分布は

$$\frac{\pi^r(1-\pi)^{n-r}}{\beta(r+1, n-r+1)}d\pi \tag{5b.3.1}$$

となる．次の試行が失敗である確率は，公式 (5b.2.9) を $\alpha=\gamma=1$，$m=1$，$s=0$ として用いれば，$(n+1-r)/(n+2)$ で与えられる．特に，n 回連続して失敗したとき，さらに失敗をする確率は，$(n+1)/(n+2)$ となる．これが有名な継続の法則である．

しかしながら (5b.3.1) を導いた手順は全く任意性に富んだものである．なぜならば，母数 π の代わりに，$\sin\theta=2\pi-1$ なる θ を母数とし，θ に一様分布を仮定すれば，r が与えられたときの θ の事後分布は

$$(1+\sin\theta)^r(1-\sin\theta)^{n-r}d\theta \tag{5b.3.2}$$

に比例するものであることがわかる．もとの母数 π の事後分布を (5b.3.2) から導けば，

$$\pi^{r-\frac{1}{2}}(1-\pi)^{n-r-\frac{1}{2}}d\pi \tag{5b.3.3}$$

に比例するが，これは (5b.3.1) とは異なったものである．このような任意性あるいは不一致性は，事前分布を指定するどんな規則においても，避けることのできないものである．

5b.4　経験的ベイズ推定の手順

F が完全にわかっていなければ，未知の変量について確率的命題を述べるために，条件つき分布 (5b.1.2) を用いることができないことは明らかである．しかし，ある場合には，θ の分布関数が完全にはわからなくても，$g(\theta)$ の $\mathbf{X}$ への回帰を計算することは可能であるかもしれない．しかしその推定量は，$\mathbf{X}$ だけの周辺分布の指標である未知母数を含んでいるかもしれない．$\mathbf{X}$ の過去のデータがあれば，$\mathbf{X}$ の周辺分布は経験的に推定でき，したがって回帰推定による1つの推定値も得られる．この手順をいくつかの例で示そう．

ただ1つの母数 θ と確率変数 $X_1,\cdots,X_n$（これを全体として記号 $\mathbf{X}$ であらわす）の同時確率密度を

$$(2\pi\sigma_1^2)^{-\frac{1}{2}}e^{-(\theta-\mu)^2/2\sigma_1^2}(2\pi\sigma_2^2)^{-n/2}\ e^{-\Sigma(X_i-\theta)^2/2\sigma_2^2} \tag{5b.4.1}$$

とする．ここで $\mu,\sigma_1^2,\sigma_2^2$ は未知である．$X_1,\cdots,X_n$ の周辺密度 [(5b.4.1) を θ について積分すれば得られる] は

$$\exp\left\{-\frac{n(\bar{X}-\mu)^2}{2(n\sigma_1^2+\sigma_2^2)}-\frac{S^2}{2\sigma_2^2}\right\} \tag{5b.4.2}$$

に比例する．ここで $\bar{X}=(X_1+\cdots+X_n)/n$，$S^2=\Sigma(X_i-\bar{X})^2$ である．$\mathbf{X}$ を与えたときの θ の条件つき密度は，

$$e^{-(\theta-\alpha)^2/2\beta^2} \tag{5b.4.3}$$

に比例する．ここで

$$\alpha=\frac{(\mu/\sigma_1^2)+(n\bar{X}/\sigma_2^2)}{(1/\sigma_1^2)+(n/\sigma_2^2)},\qquad \beta^2=\frac{\sigma_1^2\sigma_2^2}{n\sigma_1^2+\sigma_2^2}$$

である．(5b.4.3) から，$E(\theta|\mathbf{X})=\alpha$ である．つまり，θ の $\mathbf{X}$ への回帰は α である．この α は未知の定数 $\mu,\sigma_1^2,\sigma_2^2$ を含んでいるが，これらは $\mathbf{X}$ だけの周辺確率密度 (5b.4.2) の母数である．$\mathbf{X}$ について過去の観測値があるならば，これらの未知の母数はすべて，下に示すようにして推定することができる．

過去のデータは，n 個の観測値の p 個の独立な群

$$\begin{matrix} X_{11} & \cdots & X_{n1} \\ & \cdots\cdots\cdots & \\ X_{1p} & \cdots & X_{np} \end{matrix}$$

から成るものとしよう．級間および級内の分散分析を行えば，次のような，σ_1^2 と σ_2^2 を含む期待値をもつ表が得られる．なお，これは **4f.1** から **4f.3** で研究した分散成分の推定の特別な場合である．

	S.S.	期待値
級　間	B	$(p-1)(\sigma_2^2+n\sigma_1^2)$
級　内	W	$p(n-1)\sigma_2^2$

したがって σ_1^2 と σ_2^2 の推定値は

$$\hat{\sigma}_2^2=\frac{W}{p(n-1)},\qquad \hat{\sigma}_1^2=\frac{B}{n(p-1)}-\frac{W}{pn(n-1)}$$

である．μ の推定値は観測値の総平均

$$\hat{\mu}=\frac{\sum\sum X_{ij}}{pn}$$

により与えられる．それゆえ，θ の，将来の観測値 $\mathbf{X}'=(X_1,\cdots,X_n)$ への回帰の推定値は

$$\left(\frac{\hat{\mu}}{\hat{\sigma}_1^2}+\frac{n\bar{X}}{\hat{\sigma}_2^2}\right)\div\left(\frac{1}{\hat{\sigma}_1^2}+\frac{n}{\hat{\sigma}_2^2}\right)$$

で与えられる．この推定値は，観測値の平均 $\bar{X}$ と，過去のデータに基づく推定値 $\hat{\mu},\hat{\sigma}_1^2,\hat{\sigma}_2^2$ だけの関数である．

読者は，この問題が，θ と $\mathbf{X}$ 双方についての過去のデータを同時に必要とする古典的な回帰の問題とは異なることに気づかれるだろう．これは，この問題では，回帰にあらわれる母数が $\mathbf{X}$ についての観測値だけから推定可能であるからであるが，このことは常に可能であるとは限らない．

観測値が与えられたとき，ポアソン分布の母数 μ を μ の事前分布に何の仮定もおかずに推定するという Robbins (1955) によって与えられた興味深い例がある．いま，(μ,X) が確率変数であり，μ が与えられたとき X はポアソン分布に従うものとする．

x を確率変数 X の観測値とする．$X=x$ が与えられたときの μ の条件つき分布は，公式 (5b.1.2) を用いて，

$$e^{-\mu}\frac{\mu^x}{x!}dF(\mu)\div G(x)$$

で与えられる．ここで

$$G(x)=\int e^{-\mu}\frac{\mu^x}{x!}dF(\mu)$$

は x の周辺分布である．μ の X への回帰は，$X=x$ のとき

$$\int \mu e^{-\mu}\frac{\mu^x}{x!}dF(\mu) \div G(x) = (x+1)G(x+1) \div G(x)$$

となる．いま，$G(x)$ と $G(x+1)$ は未知である．しかし，確率変数 X についての過去の観測値

$$x_1, x_2, \cdots, x_N$$

があれば，$G(x)$ の1つの推定値は

$$\hat{G}(x) = \frac{\text{その値が} x \text{に等しい過去の観測値の個数}}{N}$$

で与えられ，$G(x+1)$ の推定値 $\hat{G}(x+1)$ も同様にして得られる．それゆえ，μ の回帰による1つの推定値は $(x+1)\hat{G}(x+1)/\hat{G}(x)$ である．

読者は，μ と X の観測値を同時に得ることはできなくても，μ の X への回帰は推定可能であることに気づかれるだろう．

5b.2 から **5b.4** で，θ の事前分布 F についてどの程度の知識があるかにより推定の手順がどのように変わるかを考察した．**5b.1** 節の一般的問題の特に困難な場合は，F がどの点に集中しているかがわからない一点分布の場合である．この場合は，母数が固定されているが未知であると考えられる古典的な推定理論に帰着する．

5b.5　推　測　確　率

未知の母数に関して，確率的命題を述べることが事前分布とは無関係にできないだろうかという疑問が生じるかもしれない．このような接近法は，推測確率による議論として知られる議論を用いてフィッシャーにより展開された．この理論の複雑なところまでたちいることなく，その方法を簡単に示そう．T を θ の十分統計量とし，$g(T,\theta)$ をその分布が θ に無関係であるような θ と T の関数とする．いま

$$P[g(T,\theta)<\lambda|\theta] = F(\lambda) \tag{5b.5.1}$$

とし，$g(T,\theta)$ を

$$g(T,\theta)<\lambda \Leftrightarrow \theta > h(T,\lambda) \tag{5b.5.2}$$

であるものとすれば，方程式 (5b.5.1) は

$$P[\theta > h(T,\lambda)] = F(\lambda) \tag{5b.5.3}$$

と書くことができる．2つの変数 θ と T を含む (5b.5.3) を得るまでの議論は完全に正しい．しかし，次の段階で T_0 を確率変数 T の観測値として，

$$P[\theta > h(T_0,\lambda)] = F(\lambda) \tag{5b.5.4}$$

が成立すると主張することには議論が多い．

たとえば，$\bar{x}$ を正規母集団 $N(\theta,1)$ からの大きさ n の標本の平均とすれば，関数 $y=(\bar{x}-\theta) \sim N(0,1/\sqrt{n})$ であり，

$$P\left(\bar{x}-\theta<\frac{\lambda}{\sqrt{n}}\right) = \int_{-\infty}^{\lambda/\sqrt{n}} N(y|0,1/\sqrt{n})dy = \Phi(\lambda)$$

である．特定の標本の $\bar{x}$ の値が 5.832 であるとすると，推測確率による議論によれば，

$$P\left(\theta > 5.832 - \frac{\lambda}{\sqrt{n}}\right) = \Phi(\lambda) \tag{5b.5.5}$$

であるということになるのである．

このような議論にたいして，いくつかの異論が出されている．第1に，"θ は確率変数ではなく，たとえば特定の時点での太陽と地球との間の距離のように，固定されているが未知の値を持つ" のであり，観測値から推定されるべき値であると．第2に，たとえ θ が未知の事前分布をもつ確率変数であっても，(5b.5.5) のような命題は，事前分布と無関係に常に真であることはありえないと．なぜなら，$\bar{x}$ を与えたときの事後分布が (5b.5.5) とは異なるような事前分布を見いだすことが常にできるからである．確率的命題としての (5b.5.5) の解釈は，違った方法でしなければならない．この方向へのいくつかの試みが，Dempster (1963), Fraser (1961) そして Sprott (1961) でなされている．

5b.6　ミニマックス原理

5b.1 のように $(\theta, \mathbf{X})$ が確率変数であるとし，$g(\theta)$ の推定量 $T(\mathbf{X})$ を考える．θ が与えられたときの平均2乗誤差を

$$M_T(\theta) = \underset{\mathbf{x}}{E}\{[T(\mathbf{X})-g(\theta)]^2|\theta\} \tag{5b.6.1}$$

であらわすと，これは θ が与えられたときの $\mathbf{X}$ の条件つき分布にのみ依存している．2乗誤差の代わりに，T と $g(\theta)$ との差をあらわす他の尺度を選び，その期待値を $M_T(\theta)$ とすることも可能である．この $M_T(\theta)$ は損失関数と呼んでよい．しかし，議論をわかりやすくするため，特に2乗誤差を選ぶことにする．θ の事前分布関数 F が知られていれば，全体での平均2乗誤差は，

$$M_T = \underset{\theta}{E}[M_T(\theta)] = \int M_T(\theta)dF(\theta) \tag{5b.6.2}$$

である．すでに (5b.2.3) で示したように，$M_T(\theta)$ が (5b.6.1) のように定義されているときには，T を $g(\theta)$ の $\mathbf{X}$ への回帰，つまり，$T=E_\theta[g(\theta)|\mathbf{X}]$ と選んだときに M_T は最小値をとる．したがって，F が既知のときには申し分のない解が得られる．

F がわからないときには，(5b.6.2) のような，推定量の良さを判断する積分による尺度が存在しない．そこで，2つの推定量 T, T' のどちらを選択するかということは，2つの関数 $M_T(\theta)$, $M_{T'}(\theta)$ を直接比較しなければならないときには，難かしい問題となる．もしもすべての θ と T' にたいして，$M_T(\theta) \leqslant M_{T'}(\theta)$ となるような T が存在するならば，確かに T が最良である．しかしこのようなことは一般には成立しないので，別の方法

で考えねばならない．

すべての θ にたいして $M_T(\theta) \leqslant M_{T'}(\theta)$ であり，少なくとも1つの θ の値にたいして厳密な不等号が成立するとき，T は T' よりもよいという．推定量 T は，T よりもよい推定量が存在しないとき許容的 (admissible) であるという．許容的推定量の集合は一般には非常に広く，また，この集合に属する任意の2つの推定量 T, T' にたいして，次のような空でない θ の集合 S, S' が存在する．

$$M_T(\theta) \leqslant M_{T'}(\theta),\ \theta \in S \quad \text{そして} \quad M_{T'}(\theta) \leqslant M_T(\theta),\ \theta \in S'.$$

n 回のベルヌイ試行に基づいて，2項比率 π を推定する問題を考える．r を成功の回数とすれば，(5b. 2. 8) で考えた推定量 $(r+\alpha)/(n+\beta)$ の平均2乗誤差は

$$E\left(\frac{r+\alpha}{n+\beta}-\pi\right)^2 = \frac{n\pi(1-\pi)+(\alpha-\beta\pi)^2}{(n+\beta)^2} \tag{5b. 6. 3}$$

である．特に $\alpha=\sqrt{n}/2$，$\beta=\sqrt{n}$ とすれば，推定量

$$T'=(r+\sqrt{n}/2)/(n+\sqrt{n}) \tag{5b. 6. 4}$$

の平均2乗誤差は $n/4(n+\sqrt{n})^2$ であり，π とは無関係になる．伝統的な推定量 $T=r/n$ の平均2乗誤差は $\pi(1-\pi)/n$ である．したがって，

$$\frac{n}{4(n+\sqrt{n})^2} \leqslant \frac{\pi(1-\pi)}{n}$$

であるような π の値にたいしては，つまり区間

$$\pi \in \left(\frac{1}{2}\pm a\right),\quad \text{ここで}\quad a=\frac{1}{2}\left[1-\left(1+\frac{1}{\sqrt{n}}\right)^{-2}\right]^{1/2},$$

では，T' は T よりもよい推定量である．それ以外の区間 $\left(0, \frac{1}{2}-a\right)$ と $\left(\frac{1}{2}+a, 1\right)$ では，T が T' よりもよい推定量である．しかしながら，T' の方が T よりもよい推定量であるような区間の長さは $n\to\infty$ のとき零となる．n が有限のときに，T と T' のどちらを選ぶかは，別の規準によらざるをえない．不偏性のような制約をおけば，T が平均2乗誤差最小の推定量として選ばれることはすでに示した．ここでは，T よりも T' を選ぶような別の原理について考える．

ミニマックス原理　この原理の背後にある考え方は，推定量 T のよさを $\sup_\theta M_T(\theta)$，つまり起こりうる最悪の値によって判断するということである．ミニマックス原理は $\sup_\theta M_T(\theta)$ が最小になるような推定量を選ぶことを提案する．

いま，もしも

$$\sup_\theta M_{T_1}(\theta) < \sup_\theta M_{T_2}(\theta)$$

ならば，$\sup_\theta M_T(\theta)$ という尺度に関して，T_1 は T_2 よりもよい推定量であり，もしも，

$$\sup_\theta M_{T^*}(\theta) \leqslant \sup_\theta M_T(\theta) \qquad \text{すべての } T \text{ にたいして},$$

となるような T^* が存在するならば，T^* はミニマックス推定量であるといわれる．

しかしながら，ミニマックス推定量を求める簡単な方法はない．しかし，次の結果は，それによれば，ある場合にはミニマックス推定量であることがわかるという意味で重要である．

5b. 1 のように $(\theta, \mathbf{X})$ が確率変数である場合を考え，$g(\theta)$ を $T(\mathbf{X})$ で推定するときの損失を2乗誤差 $[T(\mathbf{X})-g(\theta)]^2$ で測ることにする．いま，θ の事前分布関数 F^* が存在し，それに付随した，$\mathbf{X}$ を与えたときの g の回帰推定量 T^* ((5b. 2. 3) を参照) にたいして，損失関数

$$M_{T^*}(\theta)=\underset{\mathbf{X}}{E}\{[T^*(\mathbf{X})-g(\theta)]^2|\theta\}$$

が θ に無関係であるとする．そのとき，T^* は指定した損失関数についてミニマックス推定量である．

T^* は θ の事前分布関数 F^* に関する回帰推定量であるから，全体での平均2乗誤差

$$\int M_{T^*}(\theta)dF^*(\theta)$$

は最小である．つまり，T を任意の他の推定量とすれば

$$\int M_{T^*}(\theta)dF^*(\theta) \leqslant \int M_T(\theta)dF^*(\theta)$$

である．しかし左辺の量は定数であり，これを c とすれば，

$$c \leqslant \int M_T(\theta)dF^*(\theta)$$

であるので，任意の T にたいして $c \leqslant \sup_\theta M_T(\theta)$ となる．つまり，T^* はミニマックス推定量である．

2項比率を推定する問題では，(5b. 6. 4) で与えられる推定量 T' は，$\alpha=\sqrt{n}/2$，$\gamma=\sqrt{n}/2$ と選べば，事前確率密度 (5b. 2. 4) にたいするベイズ推定量 (5b. 2. 8) であることがわかる．さらに，その損失関数は $n/4(n+\sqrt{n})^2$ であり，π には無関係である．それゆえ

$$T^*=\frac{r+\sqrt{n}/2}{n+\sqrt{n}}$$

は，平均2乗誤差最小の規準に関する π のミニマックス推定量である．

一般の損失関数 $W[T(\mathbf{X}), g(\theta)]$ の場合には，まずベイズ推定量を決定すること，つまり

$$\int W[T(\mathbf{X}), g(\theta)]\frac{P(\mathbf{X}|\theta)}{G(\mathbf{X})}dF(\theta)$$

を最小にする関数 $T(\mathbf{X})$ を決定することが問題になる．もしも，ある事前分布 F^* が存在し，それに対応するベイズ推定量 T^* にたいして，損失関数の期待値

$$M_{T^*}(\theta)=\int W[T^*(\mathbf{X}), g(\theta)]P(\mathbf{X}|\theta)dv$$

が θ に無関係であるならば，T^* はミニマックス推定量であることがわかる．

読者は，規準が異なると解（推定量）も異なるということに驚かれるべきではない．どのような状況でも，規準の選び方は推定値を得る目的に依存するのである．たとえば，ミニマックス推定量は母数空間の広い範囲で，他の推定量よりも損失が大きいことがあるかもしれない．しかし，推定値から生じる不当に大きい損失にたいして保証を得たいという場合には有用であろう．

5b.7　不変性の原理

(5a.2.5) では，その理由を説明せずに，推定量の集まりを移行性，つまり

$$f(x_1+\alpha,\cdots,x_n+\alpha)=\alpha+f(x_1,\cdots,x_n) \qquad (5b.7.1)$$

という性質をもつものに限定した．$x_1,\cdots,x_n$ の確率密度は，$p(x_1-\theta,\cdots,x_n-\theta)$ という形であったが，これは移行のもとで不変である．つまり，α を既知の定数とし，x_i の代わりに $y_i=x_i+\alpha$ を観測値と考えれば，y_i は密度 $p(y_1-\theta',\cdots,y_n-\theta')$ をもつ，ここで $\theta'=\theta+\alpha$ である．確率密度が同じ形であるので，x_i による θ の推定関数と y_i による θ' のそれとは同じ形であるべきである．したがって，$f(x_1,\cdots,x_n)$ で θ を推定するのならば，$f(y_1,\cdots,y_n)$ で $\theta'=\theta+\alpha$ を推定することになり，2つの推定が矛盾しないためには (5b.7.1) が成り立たねばならない．すでに2章の付録 **2C** で，不変性の原理は，一般の決定問題（推定，仮説検定等を含めて）に適用できるように，より一般的な構成で定式化されている．

不変性の原理については文献で数多く論じられているが，その応用についてはいくつかの議論があるだけである［Rao (1962d) のなかのバーナードによるコメントと Stein (1962) を参照せよ］．

独立な確率変数 $X_i\sim N(\mu_i,1)$，$i=1,\cdots,n$ を考え，損失関数

$$(t_1-\mu_1)^2+\cdots+(t_n-\mu_n)^2 \qquad (5b.7.2)$$

を用いて，$\mu_1,\cdots,\mu_n$ を同時推定する問題を考える．この決定問題は，確率変数の移行のもとで不変であることがわかる．このような場合には，推定量 t_i は移行性（2章の付録 **2C** を参照せよ），つまり

$$t_i(x_1+\alpha_1,\cdots,x_n+\alpha_n)=\alpha_i+t_i(x_1,\cdots,x_n),\qquad i=1,\cdots,n \qquad (5b.7.3)$$

という性質を持つべきである．(5b.7.3) の条件のもとでは，t_i を x_i，つまり確率変数 X_i の観測値としたときに，損失(5b.7.2)の期待値は最小になる．しかしながら，別の推定量

$$t_i^*=\left(1-\frac{n-2}{\sum x_i^2}\right)x_i,\qquad i=1,2,\cdots,n \qquad (5b.7.4)$$

を考えることにする．そうすると，Stein (1962) によって

$$E\left[\sum(t_i^*-\mu_i)^2\right]=n-(n-2)^2E[(2K+n-2)^{-1}] \qquad (5b.7.5)$$

であることが示されている．ここで K は平均 $(\sum\mu_i^2)/2$ のポアソン変量である．ところが $E[\sum(x_i-\mu_i)^2]=n$ であるので，$n\geqslant 3$ のとき，これは (5b.7.5) より大きくなる．実際のところ，(5b.7.5) は，$\sum\mu_i^2=0$ のとき2という値をとり，$\sum\mu_i^2\to\infty$ のとき n に近づく．したがって，改良の効果が大きいといえるが，推定量 t_i^* は不変性の原理を満足していない．

この例は，どんな原理も，いかに心を引きつけるものであっても，普遍的妥当性は持ちえず，特定の問題において適切であるか否かは，その原理から導かれる解の正当性から判断されねばならないということを示しているだけである．

5c　大標本における推定の規準

5c.1　一　致　性

いま，$\{x_n\}$，$n=1,2,\cdots$ を，母数 θ について情報を与える観測値（確率変数）の系列とし，T_n をはじめの n 変量に基づく θ の推定量とする．ここでの目的は，$n\to\infty$ のときの系列 $\{T_n\}$ の性質を調べることである．

推定量 T_n は，T_n が θ に確率収束するか，もしくは，確率1で $T_n\to\theta$ のとき，θ の一致推定量であるという．前者を弱い一致性 W.C.(weak consistency) と呼び，後者を強い一致性 S.C. (strong consistency) と呼ぶ．

いま定義した一致性は，推定量 T_n の，$n\to\infty$ のときの極限の性質だけを問題にしている．したがって，実際の場面で，これを推定の規準として適用するには，相当の注意が必要である．なぜなら，もし T_n が一致推定量であれば，それを

$$T_n'=\begin{cases}0 & n\leqslant 10^{10}\ のとき\\ T_n & n>10^{10}\ のとき\end{cases} \qquad (5c.1.1)$$

という推定量でおきかえても，やはり一致推定量である．しかし標本の大きさが大きいが，無限には大きくないという実際の場面では，これは推定量として採用されることはないだろうからである．

それゆえ，一致性の規準は特定の標本の大きさに関しては，妥当なものではない．しかしながら，推定量がみたすべき望ましい性質の1つは，観測値の増加とともに，推定量が推定される量の真値に近づいていくことである．

たとえば，$x_1,x_2,\cdots$ を，密度関数が $[1+(x-\mu)^2]^{-1}$ に比例するコーシー母集団からの観測値とする．$\bar{x}_n=(x_1+\cdots+x_n)/n$ は，ただ1つの観測値の分布と同じ分布をする

ことが知られている（3a.4を参照せよ）．それゆえ，$|\bar{x}_n-\mu|>\delta$ である確率は，すべての n について同じであるので，$\bar{x}_n$ は μ の一致推定量ではない．しかしながら，観測値の中央値は一致推定量であることが知られている．

$\{x_n\}$ が i.i.d. なる変量の系列であり，$E(x_i)$ が存在するならば，大数の強法則によって，$\bar{x}_n$ は平均の一致推定量であることがわかる．コーシー分布の場合には，後の条件がみたされていない．

チェビシェフの不等式［(ⅲ), 2b.2］を用いることにより，もし $E(T_n)=\theta_n\to\theta$ でありかつ $V(T_n)\to 0$ ならば，T_n は θ の弱い一致推定量であることを示すのは，容易である．$x_1,\cdots,x_n$ が $N(\theta,\sigma^2)$ からの n 個の独立な観測値であるとき，σ^2 の推定量 $T_n=\sum(x_i-\bar{x})^2/n$ を考える．そのとき

$$E(T_n)=\frac{n-1}{n}\sigma^2\to\sigma^2,\qquad V(T_n)=\frac{2\sigma^4(n-1)}{n^2}\to 0$$

である．したがって，T_n は σ^2 の不偏推定量ではないが，一致推定量である．

フィッシャーの一致性 F.C.(Fisher consistency) (Fisher (1922, 1956))　フィッシャーは別の一致性の定義を導入したが，それによれば，推定関数に制約をおき，(5c.1.1) のような統計量を考慮の対象から外すことになる．その定義はどんな標本の大きさにたいしても適用でき，したがって実際の応用上意味深いものだろう．

まず，母数のベクトル θ の関数である $\pi_1(\theta),\cdots,\pi_k(\theta)$ を各群の出現確率としてもつ有限の多項分布からの標本を考える．n を全標本の大きさ，n_i を第 i 群の観測度数，$p_i=n_i/n$ を第 i 群の観測比率とし，$g(\theta)$ を推定する母数の関数とする．

推定量 T は次の2つの条件をみたすとき，$g(\theta)$ の F.C. 推定量であるという．(ⅰ) T は，$x_i\geqslant 0$ で $x_1+\cdots+x_k=1$ であるようなベクトル $(x_1,\cdots,x_k)$ の集合上で連続関数として定義され，$x_i=p_i,\ i=1,\cdots,k$ のときの T の値は，標本に基づく $g(\theta)$ の推定値である．(ⅱ) すべての可能な θ の値にたいして，$x_i=\pi_i(\theta),\ i=1,\cdots,k$ のときの T の値が $g(\theta)$ である，つまり，$T[\pi_1(\theta),\cdots,\pi_k(\theta)]\equiv g(\theta)$．

実際のところ，この定義が要求しているのは，推定量が観測比率だけの陽な関数であり，$T(p_1,\cdots,p_k)$ と書ける量であることと，観測比率がたまたま真の比率（確率）に一致したときには，推定値も $g(\theta)$ の真値と一致すること，である．実際，真の比率にたまたま気づきながら，推定の方法によっては，$g(\theta)$ の真値を推定値として得られないとしたら，奇妙なことだろう．この意味で，F.C. は方法の一致性 M.C. (method consistency) とよんでもよい．

上に定義した F.C. 推定量が S.C. 推定量でもあることは，$p_i\xrightarrow{\text{a.s.}}\pi_i(\theta)$ であり，T の連続性により［2c.4 の (xiii) を用いて］

$$T(p_1,\cdots,p_k)\xrightarrow{\text{a.s.}}T[\pi_1(\theta),\cdots,\pi_k(\theta)]=g(\theta)$$

であることから容易にわかる．

一般の場合，$x_1,\cdots,x_n$ を分布関数 $F(\cdot,\theta)$ をもつ確率変数の i.i.d. なる観測値とし，S_n を観測値に基づく経験分布関数とする．推定量 T は次の条件をみたすとき，F.C. 推定量であるという．

（a）T は分布関数の空間上で弱連続な汎関数として定義され，与えられた S_n にたいする T の値は観測値に基づく $g(\theta)$ の推定値である．

（b）$T[F(\cdot,\theta)]\equiv g(\theta)$ が θ について恒等的に成立する．

多項分布の場合のように，F.C. $\Rightarrow$ S.C. が次のようにして示せる．6章の (6f.1.3) で示されるように

$$\sup_x|S_n(x)-F(x,\theta)|\xrightarrow{\text{a.s.}}0$$

であるので，

$$T[S_n]\xrightarrow{\text{a.s.}}T[F(\cdot,\theta)]=g(\theta)$$

となる．

F.C. の概念は，i.i.d. なる観測値にだけしか適用できないことに注意する．

5c.2　有　効　性

推定量の有効性の概念については，統計の文献において，若干混乱があるようである．(5a.2.15) において，T が母数 θ の関数 $g(\theta)$ の不偏推定量であるならば，ある条件のもとで，

$$V(T)\geqslant\frac{[g'(\theta)]^2}{\mathscr{I}(\theta)}\qquad(5c.2.1)$$

であることが示された．ここで $\mathscr{I}(\theta)$ は，標本に含まれる θ についてのフィッシャーの情報である．不等式 (5c.2.1) から，いく人かの研究者は

$$V(T)=\frac{[g'(\theta)]^2}{\mathscr{I}(\theta)}$$

であるとき，つまり，(5c.2.1) の下界に到達したとき，T は $g(\theta)$ の有効（不偏）推定量であると定義している．

論理的には，これはいい定義ではないようである．なぜなら $[g'(\theta)]^2/\mathscr{I}(\theta)$ という表現は1つの下界にすぎず，多くのより鋭い下界が存在し，そのなかのどれを用いても，有効性を定義することができるからである．さらに，多くの場合に，最小の到達可能な分散は下界 (5c.2.1) より大きいことが知られている．たとえば，正規母集団からの n 個の観測値に基づく σ^2 の不偏推定量 $s^2=\sum(x_i-\bar{x})^2/(n-1)$ は，最小の分散 $2\sigma^4/(n-1)$ をもつが，下界 (5c.2.1) は $2\sigma^4/n$ という値をとるのである（(5a.3.6) を参照せよ）．有効

性を分散と結びつけたいならば，不偏推定量が，最小の到達可能な分散をもつときに有効であると定義すればよく，それが，与えられた場面でどんな推定量も到達不可能であるような特定の値を分散にもつ必要はない．しかしながら，この場合，有効性という用語は最小分散と同義語となり，推定量の真値のまわりの集中度といったような本質的な性質について触れるものではなくなるだろう．

ときには，有効性は極限の性質として定義される．推定量 T_n（大きさ n の標本に基づいている）は

$$\left.\begin{array}{l} E(T_n)\to g(\theta) \\ |V(T_n)-[g'(\theta)]^2/\mathscr{I}|\to 0\end{array}\right\}, \quad n\to\infty \text{ のとき}, \qquad (5c.2.2)$$

をみたすとき，$g(\theta)$ の漸近的に有効な推定量であるという．このように定義された有効性は，$E(T_n)$ や $V(T_n)$ は存在しないが，T_n の分布は $n\to\infty$ のとき $g(\theta)$ に集中するという場合には適用できない．

CAN 推定量　長い間使われてきており，上の定義とはやや異なる漸近的有効性の概念は，推定量の漸近分散に基づいている．推定量の集まりは，一致性をもった漸近的に正規分布に従う CAN (consistent asymptotically normal) 推定量に限定される．推定量 T_n は，$\sqrt{n}[T_n-g(\theta)]$ の漸近分布が正規であるとき，$g(\theta)$ の CAN 推定量であるという．CAN 推定量T_n は，$\sqrt{n}[T_n-g(\theta)]$ の極限分布の分散が最小可能値をもつとき，最良あるいは有効であるという．i. i. d.なる観測値を考えたときには，$\sqrt{n}[T_n-g(\theta)]$ の極限分布の分散は下界 $[g'(\theta)]^2/i$ をもつと考えられていた．ここで i は，1つの観測値に含まれる θ についてのフィッシャーの情報である．漸近分布について，上に述べた下界に到達するような推定量 T_n は，有効推定量であると考えられる．しかし残念ながら，推定量にさらに条件を置かなければ，漸近分散の下界についてのこのような結果は成立しない (Kallianpur and Rao (1955d), Rao (1963a))．事実，一致推定量の漸近分散は任意に小さくできるので，漸近分散に基づく有効性の概念は無意味なものである．

この事実の一例がホッジスによって与えられている (Le Cam (1953) を参照せよ)．T_nを $\sqrt{n}(T_n-\theta)$ の漸近分散が $v(\theta)$ であるような一致推定量とする．いま，推定量

$$T_n' = \begin{cases}\alpha T_n, & |T_n|<n^{-1/4} \text{ のとき} \\ T_n, & |T_n|\geqslant n^{-1/4} \text{ のとき}\end{cases} \qquad (5c.2.3)$$

を考える．ここでαは定数である．そのとき，$\sqrt{n}(T_n'-\theta)$ もまた漸近的に正規分布し，分散は $\theta=0$ で $\alpha^2 v(0)$，そのほかでは $v(\theta)$ となることが容易にわかる．α は任意であるから，T' の漸近分散は，$\theta=0$ では T のそれより小さく，そのほかでは等しいようにすることができる．したがって，CAN 推定量の漸近分散には意味のある下界が存在しないので，推定量にさらに条件をおかなければ，最良の CAN 推定量というものは存在しな

い．統計の文献では，最良の CAN 推定量は最良漸近正規 BAN (best asymptotically normal) 推定量とよばれている．いま示したのは，推定量の集まりに適当な制約をおかなければ，BAN 推定量は存在しないということである．後の節で，いくつかの制約のおき方について調べる．

漸近効率の Fisher (1925) の定義　いま，i. i. d. なる観測値を考え，$i(\theta)$ を1つの観測値に含まれる θ についての情報とする．(5a. 4. 3) で示した通り，T_nを θ の推定量とし，ni_{T_n} を T_n に含まれる θ についての情報とすれば，$ni_{T_n}\leqslant ni$ つまり $i_{T_n}\leqslant i$ となる．フィッシャーは，比 i_{T_n}/i（$\leqslant 1$ である）を，有限の n についての推定量 T_n の効率と定義し，$n\to\infty$ のときの i_{T_n}/i の極限を，大標本における推定量の（漸近）効率と定義した．

このように定義された効率は，母数の関数についての偏りとか一致性などのような性質について，まったく触れていないことに注意する．事実，(i_{T_n}/i) が1である，あるいは $n\to\infty$ のとき1に近づくという意味で，T_n が有効であれば，T_n と1対1の関数もまた，θ について有効である．

5a. 4 で，フィッシャーの情報が，未知の母数についての確率分布の内在的特性を測るものであることを述べた．そこで考えた内在的特性とは，接近した母数の値の間の判別力のことであった．大きさ n の標本のもつ情報は ni であり，統計量 T_n のそれは $ni_{T_n}\leqslant ni$ である．等号の成り立つときには，標本と統計量との判別力は同じである．しかし統計量はデータを縮約したものであり，したがってその判別力は全標本のそれよりは大きくはない．したがって，いくつかの統計量のなかから選択するとすれば，ni_{T_n}/ni が最大になるような統計量を選ぶことが好ましい．

T_n が θ の十分統計量であるときには，$ni_{T_n}=ni$ であるので，任意の n について情報の損失がないことはすでに示した通りである．しかし，十分統計量（1次元の可測な確率変数）は常に存在するとは限らない．しかしながら，(**5d** 節で示すように) $\lim i_{T_n}/i=1$ という性質をもつ推定量の広い集まりを見いだすことは可能である．

漸近有効性の新しい定義　$P(\cdot,\cdots,\cdot|\theta)$ を確率変数 $x_1,\cdots,x_n$ の確率密度とし，

$$z_n=\frac{1}{n}\frac{d\log P(x_1,\cdots,x_n|\theta)}{d\theta} \qquad (5c.2.4)$$

と定義する．θ の一致推定量 T_n は，

$$\sqrt{n}|T_n-\theta-\beta(\theta)z_n|\to 0 \qquad (5c.2.5)$$

が確率収束の意味か，あるいは確率1で成立するとき，1次の有効 f. o. e. (first order efficient) 推定量であるという．ここで β は観測値を含まない (Rao (1961a, 1961b, 1962d, g, 1963a, 1965a))．

条件 (5c. 2. 5) より，T_n と z_n の漸近相関が1であるので，任意の推定量の1次の有効性は，T_n と z_n の（漸近）相関の2乗により測ることができる．

母数が複数個ある場合には，導関数のベクトル

$$\mathbf{Z}_n' = (z_n^1, \cdots, z_n^q)$$

$$z_n^i = \frac{1}{n}\frac{\partial \log P}{\partial \theta_i}, \qquad i = 1, \cdots, q$$

と，推定量の真値からの差のベクトル

$$\mathbf{D}_n' = (\mathbf{T}_n - \theta)' = (T_n^1 - \theta_1, \cdots, T_n^q - \theta_q)$$

を用いて，

$$\sqrt{n}\,|\mathbf{D}_n - \mathbf{B}\mathbf{Z}_n| \to 0 \tag{5c. 2. 6}$$

が確率収束の意味か，あるいは確率1で成立するとき，$\mathbf{T}_n$ は f. o. e. であるという．ここで $\mathbf{B}$ は定数行列であり θ には依存するかもしれない．

漸近有効性の定義 (5c. 2. 5) および (5c. 2. 6) から，以下の結果が得られる．

（i） $x_1, x_2, \cdots$ を，確率密度 $p(x, \theta)$ をもつ，i. i. d. なる確率変数の列とする．さらに

$$\int p'(x,\theta)dx = 0 \quad \text{かつ} \quad \int \frac{p'^2}{p}dx = i(\theta) > 0 \tag{5c. 2. 7}$$

であるとする．そうすると，条件 (5c. 2. 5) が成り立てば，$\sqrt{n}(T_n - \theta)$ の漸近分布は $N(0, \beta^2 i)$ である．

条件 (5c. 2. 7) のもとでは，中心極限定理によって，

$$\sqrt{n}\,z_n = \left[\frac{p'(x_1,\theta)}{p(x_1,\theta)} + \cdots + \frac{p'(x_n,\theta)}{p(x_n,\theta)}\right] \div \sqrt{n}$$

の漸近分布は $N[0, i(\theta)]$ である．(5c. 2. 5) より，[(ix), **2c. 4**] を用いて $\sqrt{n}(T_n - \theta)$ の漸近分布は $\beta\sqrt{n}\,z_n$ のそれと同じである．したがって証明できた．

（ii） T_n と T_n' が双方とも f. o. e. であれば，

$$\sqrt{n}\,|T_n - \alpha(\theta) - \gamma(\theta)T_n'| \to 0 \tag{5c. 2. 8}$$

が確率収束の意味で，あるいは確率1で成立するような $\gamma(\theta)$ と $\alpha(\theta) = \theta[1 - \gamma(\theta)]$ が存在するという意味で，2つの推定量は同等である．つまり，漸近的には一方は他方の線形関数である．

（iii） 母数が複数個ある場合，(i) の場合と同様の条件のもとで，$\sqrt{n}\mathbf{Z}_n$ の漸近分布は平均が零，分散共分散行列が情報行列 $\mathscr{I} = (i_{rs})$ の q 変量正規分布である．条件 (5c. 2. 6) から，$\sqrt{n}(\mathbf{T}_n - \theta)$ もまた漸近的に，分散共分散行列が $\mathbf{B}\mathscr{I}\mathbf{B}'$ の q 変量正規分布に従う．

$\sqrt{n}\mathbf{Z}_n$ の漸近正規性は，多変量の中心極限定理 [(iv), **2c. 5**] を適用すれば導かれる．したがって，残りも導かれる．

（iv） $\sqrt{n}\,|T_n - \theta - \beta z_n|$ が零に確率収束すれば，$n \to \infty$ のとき $\lim(i_{T_n}/i) \to 1$ で

ある．

(iv)の結果は，有効性の定義 (5c. 2. 5)から，フィッシャーの有効性が導かれることを示している．この結果の証明はやや複雑である (Doob (1934, 1936), Rao(1961a)を参照せよ).

後の節で，定義 (5c. 2. 5) あるいは (5c. 2. 6) をみたす推定量の集合が空ではないことが示される．したがって，漸近有効性の定義 (5c. 2. 5) と (5c. 2. 6) は満足のいくもののようである．

最良 CUAN 推定量　CAN 推定量を考えることの1つの利点は，推定量 T_n に基づく θ についての推論が，正規分布を用いて導かれることである．したがって $\sqrt{n}(T_n - \theta)$ の漸近分布が $N[0, v(\theta)]$ であれば，仮説 $\theta = \theta_0$ を検定するためには，棄却域

$$\frac{\sqrt{n}\,|(T_n - \theta_0)|}{\sqrt{v(\theta_0)}} > d_\alpha \tag{5c. 2. 9}$$

を用いることができる．ここで d_α は $N(0, 1)$ の両側確率 α の点である．$v(\theta)$ が θ について連続であれば，検定 (5c. 2. 9) は，**2c. 4** の (x, b) により

$$\frac{\sqrt{n}\,|(T_n - \theta_0)|}{\sqrt{v(T_n)}} > d_\alpha \tag{5c. 2. 10}$$

に同等である．

この検定の形から，θ についての 1 − α 信頼区間を，

$$T_n \pm \frac{d_\alpha\sqrt{v(T_n)}}{\sqrt{n}} \tag{5c. 2. 11}$$

という形で作りたくなるかもしれない．しかしながら，$\sqrt{n}(T_n - \theta)$ の分布の正規分布への収束が，θ のコンパクトな区間において一様でなければ，手順 (5c. 2. 11) は正しくない．したがって，(5c. 2. 9) と (5c. 2. 11) のような型の推論に，つまり，単純仮説の検定と信頼区間の設定の双方のために推定量を用いたのならば，推定量が一致性をもち，かつ一様に漸近的に正規分布に従う CUAN (consistent and uniformly asymptotically normal) ことが必要である．i. i. d. なる観測値の確率密度関数に適当な正則条件を設けると，T_n が CUAN であれば，その漸近分散 $v(\theta)$ は下界 $1/i(\theta)$ をもつということが示されている (Rao (1963a))．したがって，最良 CUAN 推定量は，その漸近分散 $v(\theta)$ が $1/i(\theta)$ となるような推定量である．それゆえ，CAN 推定量に，さらに正規分布へ一様収束するという条件を付加すれば，最小分散の概念は無意味ではない (Rao (1965a))．

5d　大標本における推定の方法

5d. 1　モーメント法

いま，$x_1, \cdots, x_n$ を，未知母数 $\theta_1, \cdots, \theta_q$ をもつ分布からの i. i. d. なる観測値とする．さ

らに，分布の原点まわりのモーメント（積率）がはじめの q 個は存在し，未知母数の陽な関数 $\alpha_r(\theta_1, \cdots, \theta_q)$，$r=1, \cdots, q$ として書けるものとする．いま，

$$a_r = (\sum x_i^r) \div n$$

でモーメント関数をあらわすとすれば，モーメント法とは，標本の実現値 a_{r0} と分布のモーメントとを等しいとおき，すなわち

$$\alpha_r(\theta_1, \cdots, \theta_q) = a_{r0}, \qquad r = 1, \cdots, q \tag{5d.1.1}$$

とし，これを $\theta_1, \cdots, \theta_q$ について解くという方法である．

a_r は n 個の確率変数の平均であるので，r 次の原点まわりのモーメント $E(x_i^r)$ が存在すれば，大数の法則により確率1で $a_r \to \alpha_r(\theta_1, \cdots, \theta_q)$ が成り立つ．したがって，a_r は α_r の（不偏であるばかりでなく）一致推定量であることがわかる．もし $\theta_1, \cdots, \theta_q$ と $\alpha_1, \cdots, \alpha_q$ の間の対応が1対1であり，逆関数

$$\theta_i = f_i(\alpha_1, \cdots, \alpha_q), \qquad i = 1, \cdots, q$$

が $\alpha_1, \cdots, \alpha_q$ について連続であれば，

$$\hat{\theta}_i = f_i(a_{10}, \cdots, a_{q0}), \qquad i = 1, \cdots, q$$

は (5d.1.1) の解であり，$f_i(a_1, \cdots, a_q)$ は θ_i の一致推定量である，$i=1, \cdots, q$，ことは容易にわかる．

それゆえ，モーメント法は，適当な条件のもとでは一致推定量を与え，推定方程式は多くの場合簡単なものである．しかし，この方法は，コーシー分布の場合のように理論的な積率が存在しないときには，適用することができない．

また，この方法で得られる推定量は一般には有効性をもたない．効率の損失に関するいくつかの計算については，読者は Fisher (1922) を参照されたい．

5d.2　最小カイ2乗法およびそれに関連した方法

これらの方法は，観測値が連続な観測値であるが，適当なクラスの区間にグループ分けされている場合か，または観測値自身が有限個の互いに排反な事象の度数である場合に適用できる．どちらの場合でも，q 個の未知母数の関数である仮説的確率 $\pi_1, \cdots, \pi_k$ をもった k 群について，観測度数 $n_1, \cdots, n_k$ が得られている．まず，観測度数 $n_1, \cdots, n_k$ と仮説のもとでの期待値 $n\pi_1(\theta), \cdots, n\pi_k(\theta)$ との間の食い違いの尺度を定義する．ここで $n=\sum n_i$ である．推定値は，この尺度を θ について最小にすることにより得られる．母数の推定のために用いられる尺度のいくつかを以下に示す．

（a）　**カイ2乗**

$$\chi^2 = \sum \frac{[n_i - n\pi_i(\theta)]^2}{n\pi_i(\theta)}.$$

（b）　**修正カイ2乗**

$$\mathrm{mod}\,\chi^2 = \sum \frac{[n_i - n\pi_i(\theta)]^2}{n_i}$$

ここで，n_i が零のときは，1でおきかえるものとする．

（c）　**ヘリンガーの距離**

$$\mathrm{H.\,D.} = \cos^{-1} \sum \sqrt{(n_i/n)\pi_i(\theta)}.$$

（d）　**カルバック＝ライブラーの分離度**

$$\mathrm{K.\,L.\,S.} = \sum \pi_i(\theta) \log \frac{\pi_i(\theta)}{n_i/n}.$$

（e）　**ホールデンの不一致度**

$$D_k = \frac{(n+k)!}{n!} \sum \frac{n_i!\,\pi_i^{k+1}(\theta)}{(n_i+k)!}, \qquad k \neq -1$$

$$D_{-1} = -\frac{1}{n} \sum n_i \log \pi_i(\theta).$$

これらの方法によって得られる推定量は，適当な正則条件のもとで，一致性をもち，かつ規準 (5c.2.5) と (5c.2.6) によれば1次の有効性をもつという意味で，かなり良い推定量である．これらの方法の詳しい研究については，読者は，Neyman (1949) と Rao (1955c, 1961a, b, 1963a) の論文を参照されたい．

しかしながら，2次の有効性とよばれる性質を考慮すると，これらの方法の間には差があることが明らかになる．2次の有効性を問題にすれば，ある正則条件のもとで，この節で考えたすべての方法よりもすぐれた推定方法が存在することが示される (Rao (1961 a, b, 1962 d) を参照せよ)．それは，最尤法であり，それについてこの章の残りで詳細に議論しよう．

5d.3　最 尤 法

X が観測値の集合の実現値をあらわし，$P(X, \theta)$ が同時密度をあらわすものとする．ここで $\theta = (\theta_1, \cdots, \theta_q)$ は母数のベクトルであり集合 $\Theta \subset E_q$ に属するものとする．観測値が与えられたときの θ の尤度は，θ の関数として

$$L(\theta | X) \propto P(X, \theta) \tag{5d.3.1}$$

で定義される．最尤 m.l. (maximum likelihood) 法の原理は，θ の推定値として

$$L(\hat{\theta} | X) = \sup_{\theta \in \Theta} L(\theta | X) \tag{5d.3.2}$$

なる $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_q)$ を採用することである．上限は実現されるとは限らないので，その場合には，"m.l. に近い" 推定値として

$$L(\theta^* | X) \geqslant c \sup_{\theta \in \Theta} L(\theta | X)$$

なる θ* を得ることは可能であろう．ここで c は $0<c<1$ なる固定された数である．$\hat{\theta}$ あるいは θ* が存在しないような標本の集合もあるだろう．しかし，$P(X,\theta)$ に正則条件をおけば，このような標本の出現頻度は無視できるものであることが示される．

実際には，$l(\theta|X)=\log L(\theta|X)$ を扱う方が便利であり，その場合には (5d.3.2) の $\hat{\theta}$ は方程式

$$l(\hat{\theta}|X)=\sup_{\theta\in\Theta} l(\theta|X) \tag{5d.3.3}$$

を満足する．(5d.3.3) の上限が Θ の内点で実現され，$l(\theta|X)$ が θ の微分可能な関数であれば，その点での偏微分は零になるので，$\hat{\theta}$ は方程式

$$\frac{\partial l(\theta|X)}{\partial\theta_i}=0,\quad i=1,\cdots,q \tag{5d.3.4}$$

の解である．方程式 (5d.3.4) は m.l. 方程式とよばれ，その任意の解は m.l. 方程式推定値とよばれる．

観測値 X の標本空間の上で (5d.3.3) により定義される関数 $\hat{\theta}$ は m.l. 推定量とよばれる．m.l. 推定量についての重要な結果が，**5e** と **5f** に述べられている．

一般的な結果を議論する前に，m.l. 法が広く適用可能であることを示すため，いくつかの簡単な例を考えよう．$x_1,\cdots,x_n$ を $N(\mu,\sigma^2)$, $-\infty<\mu<\infty$, $0<\sigma^2<\infty$ からの観測値とする．そのとき

$$l(\mu,\sigma^2|X)=-n\log\sigma-\frac{\sum(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial l}{\partial\mu}=\frac{\sum x_i-n\mu}{\sigma^2}=0,\quad \frac{\partial l}{\partial\sigma^2}=\frac{\sum(x_i-\mu)^2}{2\sigma^4}-\frac{n}{2\sigma^2}=0$$

であるから，推定値

$$\hat{\mu}=\bar{x}\quad そして\quad \hat{\sigma}^2=\frac{\sum(x_i-\bar{x})^2}{n}=s^2 \tag{5d.3.5}$$

が得られる．いま

$$l(\hat{\mu},\hat{\sigma}|X)=-n\log s-\frac{n}{2},$$

$$l(\mu,\sigma|X)=-n\log\sigma-\frac{ns^2+n(\bar{x}-\mu)^2}{2\sigma^2}$$

である．$\hat{\mu},\hat{\sigma}$ が実際に尤度の上限を与えることを示すため，不等式 $l(\mu,\sigma|X)\leqslant l(\hat{\mu},\hat{\sigma}|X)$ つまり，

$$-n\log\sigma-\frac{ns^2+n(\bar{x}-\mu)^2}{2\sigma^2}\leqslant -n\log s-\frac{n}{2}$$

$$0\leqslant\left\{\left(\frac{s^2}{2\sigma^2}-\frac{1}{2}\right)-\log\frac{s}{\sigma}\right\}+\frac{(\bar{x}-\mu)^2}{2\sigma^2} \tag{5d.3.6}$$

が成立するかどうかを調べてみる．ところで，任意の $x>0$ にたいし，$\log x\leqslant(x^2-1)/2$ であるから，(5d.3.6) の括弧のなかにある第1項は非負である．第2項も非負であるから，不等式 (5d.3.6) が成り立つ．

$x_1,\cdots,x_n$ を，母数 μ をもつポアソン分布からの観測値とする．そのとき

$$l(\mu|x_1,\cdots,x_n)=-n\mu+(x_1+\cdots+x_n)\log\mu$$

であるから，推定値 $\hat{\mu}=(x_1+\cdots+x_n)/n$ が得られるが，これが尤度の上限を与えることも示される．

上の2例では，m.l. 推定量は最小十分統計量の関数であることがわかる．このことは，密度関数の2つの部分への分解——一方は十分統計量と母数の関数，他方は θ には無関係な部分に分けられることから，一般に成立する．尤度を最大にすることは，第1の因子を最大にすることと同等であるので，θ は十分統計量の関数になるのである．

尤度原理の限界　いま，1から N までの番号をつけた N 個のチップがあり，各々に未知の値 $X_1,\cdots,X_N$ が書かれているものとしよう．n 個のチップを無作為に非復元抽出で抜き出し，各チップについて2つの値

$$(a=チップの番号,\ b=そのチップの\ X\ の値)$$

を記録する．そのとき標本の結果を

$$S:[(a_1,b_1),\cdots,(a_n,b_n)]$$

とあらわすことができる．S が与えられたとき，未知母数 $\tau=X_1+\cdots+X_N$ をどのように推定したらよいだろうか？

S の確率は，観測値がどんな値であっても，$n!(N-n)!/N!$ であるから，S が与えられたときの，未知母数 $X_1,\cdots,X_N$ の尤度は，その値が，抜き出されたチップに観測された値と矛盾する値であるならば0，その他の場合には観測されないチップの値の如何に関わらず一定値であることがわかる．したがって，ゴダンベの指摘した通り，尤度原理によっては，観測されなかったチップについて，どんな母数の値の集合も，他に較べて好ましいものとして選ぶことができない（つまりいいかえれば，この原理では，未知母数の推定は不可能なのである）．

尤度は未知母数について情報を与えるものではなかったが，標本が τ についての情報をもたないとは言えない．たとえば，推定量

$$N\frac{b_1+\cdots+'_n}{n}$$

は τ の不偏推定量であり，このことは標本は τ についての情報をもっていることを示している．それゆえ，尤度による接近法には何か限界があるように思われる．

この問題に関する議論については，読者は Rao (1971b) の論文と Godambe and Sprott (1971) にある論文を参照されたい．

5e　多項分布の推定

5e.1　確率が母数に依存しない場合

多項分布は，k 個の互いに排反な事象の確率を並べたベクトル $\pi=(\pi_1,\cdots,\pi_k)$ によって指定される．いま，n 回の独立な試行の結果，第1種の事象が n_1 回，第2種の事象が n_2 回，というふうに起こるとすると，$\pi_1,\cdots,\pi_k$ が与えられたとき，$n_1,\cdots,n_k$ の起こる確率は，

$$\frac{n!}{n_1!\cdots n_k!}\pi_1^{n_1}\cdots\pi_k^{n_k} \tag{5e.1.1}$$

である．式(5e.1.1)を $L(\pi|n_1,\cdots,n_k)$，あるいは π の関数として簡単に $L(\pi)$ とあらわすことにする．どんな特定の場合にもあてはまる真の分布 π は未知ではあるが，集合 A（許容集合）に属するものとする．目的は，n 個の独立な観測値に基づいて，真の分布を推定することである．

実際には，通常，分布を定める指標として用いられるいくつかの母数をまず推定することにより，分布が推定される．理論上は，分布の推定と，分布を定める母数の推定とを区別する必要はない．だが，ある場合には，母数の推定が主要な関心事であることもあるだろう．このような問題は **5e.2** で考えよう．この節では，π つまり全体としての分布の m.l. 推定量に関する，いくつかの重要な結果を証明しよう．同様の結果が，最小カイ2乗やヘリンガーの距離最小などの推定方法についても成立する．読者は，これらの方法についての証明を同様のやり方で，簡単に与えることができるだろう．

いま，π^* が

$$L(\pi^*)\geqslant c\sup_{\pi\in A}L(\pi),\qquad 0<c<1 \tag{5e.1.2}$$

をみたすとき，これを π の近似 m.l. 推定量であると定義する．近似 m.l. 推定量（これは常に存在する）を定義するのは，

$$L(\hat{\pi})=\sup_{\pi\in A}L(\pi) \tag{5e.1.3}$$

をみたす m.l. 推定量 $\hat{\pi}$ が存在しない場合を含めて考えるためである．分布の推定量に関して次の結果が得られる．

（i）π^* を π の近似 m.l. 推定量とし，π^0 を真の分布とする．そうすると，許容集合 A について何の仮定（制約）もおかなくとも，確率1で $\pi^*\to\pi^0$ が成り立つ．

m.l. 推定量が存在するときはそれを含めて，(5e.1.2) をみたす推定量 π^* は多数あるかもしれない．結果(i)は，任意の選び方にたいして（そして，実際には，すべての選び方にたいして一様に）π^* が収束することを主張している．要素 $p_1,\cdots,p_k$ をもつベクトルを $\mathbf{p}$ であらわすことにする．

$p_i=n_i/n$ を，第 i 種の事象の観測比率とする．大数の強法則により，確率1で $p_i\to\pi_i^0$, $i=1,\cdots,k$ であるから，上に述べた結果を証明するためには，$\mathbf{p}\to\pi^0$ かつ $n\to\infty$ のとき，$\mathbf{p}$ と n の関数としての π^* が π^0 に近づくことを示せばよい．一般性を失うことなく，$\pi_i^0>0$, $i=1,\cdots,k$ とする．(5e.1.2) の対数をとり，n で割れば

$$\begin{aligned}\sum p_i\log\pi_i^*&\geqslant\frac{\log c}{n}+\sup_{\pi\in A}\sum p_i\log\pi_i\\&\geqslant\frac{\log c}{n}+\sum p_i\log\pi_i^0\end{aligned} \tag{5e.1.4}$$

を得る．一方，情報理論における初等的な不等式 (1e.6.1) により

$$\sum p_i\log\pi_i^*\leqslant\sum p_i\log p_i \tag{5e.1.5}$$

を得る．(5e.1.4) と (5e.1.5) を合わせれば，

$$\sum p_i\log p_i\geqslant\sum p_i\log\pi_i^*\geqslant\frac{\log c}{n}+\sum p_i\log\pi_i^0$$

$$0\geqslant\sum p_i\log\frac{\pi_i^*}{p_i}\geqslant\frac{\log c}{n}+\sum p_i\log\frac{\pi_i^0}{p_i}$$

であることがわかる．$\mathbf{p}\to\pi^0$ かつ $n\to\infty$ のときには，最後の項は零に近づく．それゆえ，確率変数 $p_1,\cdots,p_k$ の関数として，

$$\lim_{n\to\infty}\sum p_i\log\frac{\pi_i^*}{p_i}=0,\qquad\text{確率1で} \tag{5e.1.6}$$

が成立する．(5e.1.6) の各項の対数に，引数の値が1の近くでテイラー展開をほどこせば，

$$\begin{aligned}-\sum p_i\log\frac{\pi_i^*}{p_i}&=\sum p_i\left(\frac{\pi_i^*}{p_i}-1\right)^2\frac{1}{2y_i^2},\qquad y_i\in\left(1,\frac{\pi_i^*}{p_i}\right)\\&=\sum p_i\frac{(\pi_i^*-p_i)^2}{2z_i^2},\qquad z_i\in(p_i,\pi_i^*)\\&\geqslant\sum p_i(\pi_i^*-p_i)^2/2,\qquad z_i\leqslant 1\text{ だから,}\end{aligned}$$

である．(5e.1.6)から，確率1で $\sum p_i(\pi_i^*-p_i)^2\to 0$ となる．そのとき，p_i は確率1で π_i^0 に収束するので，π_i^* も確率1で π_i^0 に収束する．なぜならば，もしそうでなければ矛盾が生じるからである．

（ii）$A_\varepsilon=\{\pi:\sum\pi_i^0\log(\pi_i/\pi_i^0)\geqslant-\varepsilon\}$ とし，すべての十分小さい ε にたいして，A_ε は A の内部にあるものとする．そうすると，確率1で m.l. 推定量が存在する．

集合 $A_\varepsilon^c=A-A_\varepsilon$ を考えれば，そこでは

$$\sum\pi_i^0\log\pi_i<-\varepsilon+\sum\pi_i^0\log\pi_i^0$$

が成立する．S を，A_ε^c に属し，かつ

$$\sum(\pi_i^0-\varepsilon_1)\log\pi_i<-\varepsilon+\sum\pi_i^0\log\pi_i^0$$

なる π の集合とする．ここで $\varepsilon_1>0$ である．$p_i\to\pi_i^0$ であるので，$p_i>\pi_i^0-\varepsilon_1$ とできるが，その場合

$$\sum p_i\log\pi_i\leqslant\sum(\pi_i^0-\varepsilon_1)\log\pi_i<-\varepsilon+\sum\pi_i^0\log\pi_i^0$$
$$\leqslant-\varepsilon_2+\sum p_i\log\pi_i^0$$

である．ここで $\varepsilon_2\leqslant\varepsilon$ である．S の外でかつ A_ε^c に含まれる集合では，$\log\pi_i$ は有界であり，したがって $p_i\to\pi_i^0$ を考慮すれば，

$$\sum p_i\log\frac{\pi_i}{\pi_i^0}\leqslant-\varepsilon_2$$

である．それゆえ，A_ε^c では，$\mathbf{p}$ が π^0 に十分近いとき

$$n\sum p_i\log\frac{\pi_i}{\pi_i^0}\leqslant-n\varepsilon_2$$

である．つまり指数をとれば

$$L(\pi^0)\geqslant e^{n\varepsilon_2}L(\pi)\tag{5e.1.7}$$

であることがわかる．したがって

$$L(\pi_0)\geqslant\sup_{\pi\in A_\varepsilon^c}L(\pi)$$

が得られるので，確率1で

$$\sup_{\pi\in A}L(\pi)=\sup_{\pi\in A_\varepsilon}L(\pi),\qquad \pi^0\in A_\varepsilon\ \text{であるから}\tag{5e.1.8}$$

である．しかし，$L(\pi)$ は π の連続関数であり，十分小さな ε にたいして，A_ε は A の内部に属する．それゆえ，上限は A_ε のなかで実現される．

したがって，$n\to\infty$ のとき，π の m.l. 推定量は存在し，結果 (i) によって，真の分布に収束する．

(iii) λ を A の上の事前確率測度とし，A_ε を (ii) で定義された集合とする．さらに，$P(A_\varepsilon|n_i)$ を，$n_1,\cdots,n_k$ が与えられたときの事象 $\pi\in A_\varepsilon$ の事後確率とする．つまり

$$P(A_\varepsilon|n_i)=\frac{\int_{A_\varepsilon}L(\pi)d\lambda}{\int_A L(\pi)d\lambda}.$$

そうすると，

$$P(A_{\varepsilon'})=\int_{A_{\varepsilon'}}d\lambda>0$$

となる $\varepsilon'<\varepsilon$ が存在すれば，真の分布 π^0 について，$n\to\infty$ のとき確率1で $P(A_\varepsilon|n_i)\to1$ である．

結果 $P(A_\varepsilon|n_i)\to1$ より，事後確率分布は，$n\to\infty$ のとき確率1で π^0 に集中する．(i) および (ii) と同様に，$P(A_\varepsilon|n_i)$ が $\mathbf{p}$ と n の関数であることに注意し，$\mathbf{p}\to\pi^0$ かつ $n\to\infty$ のとき $P(A_\varepsilon|n_i)\to1$ であることを示せば，証明すべき確率の収束が示される．

$A_{\varepsilon'}$ のなかでは，

$$\sum\pi_i^0\log\frac{\pi_i}{\pi_i^0}\geqslant-\varepsilon'$$

であり，$\log\pi_i$ は有界である．それゆえ，$\mathbf{p}\to\pi^0$ であるから

$$\sum p_i\log\frac{\pi_i}{\pi_i^0}\geqslant-\varepsilon_1,\qquad \varepsilon'<\varepsilon_1<\varepsilon$$

である．n をかけ指数をとれば

$$L(\pi)\geqslant e^{-n\varepsilon_1}L(\pi_0),\qquad A_{\varepsilon'}\ \text{のなかでは},$$

であることがわかる．(ii) の (5e.1.7) 式において，ε_2 を $\varepsilon_1<\varepsilon_2<\varepsilon$ となるように選べるので，A_ε^c のなかでは

$$L(\pi)\leqslant e^{-n\varepsilon_2}L(\pi_0)$$

である．それゆえ

$$\int_{A_\varepsilon^c}L(\pi)d\lambda\leqslant e^{-n\varepsilon_2}L(\pi_0)[1-P(A_{\varepsilon'})]\tag{5e.1.9}$$

を得る．ところが

$$\int_A L(\pi)d\lambda\geqslant\int_{A_{\varepsilon'}}L(\pi)d\lambda\geqslant e^{-n\varepsilon_1}L(\pi_0)P(A_{\varepsilon'})\tag{5e.1.10}$$

である．(5e.1.9) と (5e.1.10) との比をとれば

$$1-P(A_\varepsilon|n_i)\leqslant e^{-n(\varepsilon_2-\varepsilon_1)}\frac{1-P(A_{\varepsilon'})}{P(A_{\varepsilon'})}$$
$$\to0,\qquad n\to\infty\ \text{のとき},$$

であるので，証明された．

5e.2　確率が母数の関数である場合

5e.1 で述べたような結果は，分布を指定する母数の推定量と真値に関しては，さらに仮定をおかなければ，成立しない．分布を定める母数が，性質のよい，分布の汎関数でないときには，困難が生じる．π_i を，許容集合 Θ に属するベクトル値母数 $\theta=(\theta_1,\cdots,\theta_q)$ の関数であると考えよう．真値 θ^0 は Θ の内点であるとする．そこで，次の一連の仮定をおく．

仮定 1.1.　$\delta>0$ が与えられたとき，

$$\inf_{|\theta-\theta^0|>\delta}\sum\pi_i(\theta^0)\log\frac{\pi_i(\theta^0)}{\pi_i(\theta)}\geqslant\varepsilon\tag{5e.2.1}$$

であるような ε が存在する．ここで $|\theta-\theta^0|$ は θ と θ^0 との間の距離である．

仮定 1.1 は，強い識別可能性の条件と呼んでよい．仮定 1.1 より，球 $|\theta-\theta^0|\leqslant\delta$ の外では，$r\to\infty$ のとき $\pi(\theta_r)\to\pi(\theta^0)$ となるような点 θ_r の系列は存在しない．つまり，θ^0 から離れていながら $\pi_i(\theta^0)$ とほとんど等しい確率の組を与えるような θ の値は存在し

ないことがわかる．

仮定 1.2. $\theta \neq \beta$ のとき，少なくとも1つの i について，$\pi_i(\theta) \neq \pi_i(\beta)$ である．これはより弱い識別可能性の条件である．

仮定 2.1. 関数 $\pi_i(\theta)$, $i=1,\cdots,k$ は θ について連続である．

仮定 2.2. 関数 $\pi_i(\theta)$ は1次の偏導関数をもつ．

仮定 2.3. 関数 $\pi_i(\theta)$ は1次の偏導関数をもち，偏導関数は θ^0 で連続である．

仮定 2.4. 関数 $\pi_i(\theta)$ は θ^0 で全微分可能である．

仮定 3. 情報行列 (i_{rs}) は θ^0 で正則である．ここで

$$i_{rs}=\sum_j \frac{1}{\pi_j}\frac{\partial \pi_j}{\partial \theta_r}\frac{\partial \pi_j}{\partial \theta_s}$$

である．

5e.1 と同様に，θ^* が

$$L[\pi(\theta^*)] \geqslant c \sup_{\theta\in\Theta} L[\pi(\theta)] \tag{5e.2.2}$$

をみたすとき，近似 m.l. 推定量であると定義する．

（i） 仮定 1.1 が成り立てば，$n \to \infty$ のとき確率1で $\theta^* \to \theta^0$ である．

5e.1 の (i) で ((5e.1.6) 式) 示したように

$$\sum p_i \log \frac{\pi_i(\theta^*)}{p_i} \xrightarrow{\text{a.s.}} 0$$

であるので，$\mathbf{p} \to \boldsymbol{\pi}^0$ のとき $\log \pi_i(\theta^*)$ が有界であることから

$$\sum \pi_i(\theta^0) \log \frac{\pi_i(\theta^*)}{\pi_i(\theta^0)} \xrightarrow{\text{a.s.}} 0 \tag{5e.2.3}$$

が得られる．それゆえ，仮定 1.1 から，確率1で $|\theta^*-\theta^0| \leqslant \delta$ である．いま δ は任意であるから，確率1で $\theta^* \to \theta^0$ である．

（ii） 仮定 1.1 および 2.1 が成り立てば，確率1で，θ の m.l. 推定量が存在し θ^0 に収束する．

(5e.1.8) より，$\mathbf{p} \to \boldsymbol{\pi}^0$ のとき

$$\sup_{\pi\in A} L(\pi) = \sup_{\pi \in A_\varepsilon} L(\pi)$$

を得るが，これを母数を使ってあらわせば，

$$\sup_{\theta\in\Theta} L[\pi(\theta)] = \sup_{|\theta-\theta^0|\leqslant\delta} L[\pi(\theta)]$$

となる．$L[\pi(\theta)]$ は θ について連続であるから，上限は閉球 $|\theta-\theta^0| \leqslant \delta$ のなかで到達される．m.l. 推定量を $\hat{\theta}$ であらわす．

（iii） 仮定 1.1 および 2.2 が成り立てば，確率1で，m.l. 推定量は方程式

$$\frac{\partial L}{\partial \theta_i}=0, \qquad i=1,\cdots,k \tag{5e.2.4}$$

の根として得られる．

δ を，球 $|\theta-\theta^0| \leqslant \delta$ が Θ の内部にはいるように選べば，$L[\pi(\theta)]$ の上限は，ある開区間のなかの点（これを $\hat{\theta}$ とする）で到達される．そうすると，$L[\pi(\theta)]$ が微分可能ならば，偏導関数の値は $\hat{\theta}$ で零でなければならない．

（iv） 仮定 1.2, 2.3 および 3 が成り立つものとする．さらに，(i^{rs}) を (i_{rs}) の逆行列とし，

$$z_r=\sum_{i=1}^k \frac{p_i}{\pi_i(\theta^0)}\frac{\partial \pi_i}{\partial \theta_r^0}$$

とする．そうすると，尤度方程式 (5e.2.4) の一致性をもつ根 $\bar{\theta}$ が存在し，それは m.l. 推定量ではないかもしれないが，

$$\sqrt{n}\,|(\bar{\theta}_r-\theta_r^0)-i^{r1}z_1-\cdots-i^{rq}z_q| \xrightarrow{P} 0, \qquad r=1,\cdots,q \tag{5e.2.5}$$

が成立する．つまり，m.l. 方程式推定量は (5c.2.6) の意味で f.o.e. であり，その漸近分布は q 変量正規分布である．

半径 δ，中心 θ^0 の球 $|\theta-\theta^0| \leqslant \delta$ と，その球面上の関数

$$\sum \pi_i(\theta^0) \log \frac{\pi_i(\theta^0)}{\pi_i(\theta)} \tag{5e.2.6}$$

を考える．$\pi_i(\theta)$, $i=1,\cdots,k$ は連続であるから，(5e.2.6) の下限は球面上で到達され，仮定 1.2 から，下限はある正の値 ε 以下ではない．したがって

$$\inf_{|\theta-\theta^0|=\delta} \sum \pi_i(\theta^0) \log \frac{\pi_i(\theta^0)}{\pi_i(\theta)} > \varepsilon > 0 \tag{5e.2.7}$$

である．さらに，〔δ を十分小さくとれば〕$\log \pi_i(\theta)$ は球面上で有界であり，$\mathbf{p}$ が $\pi(\theta^0)$ に十分近いときには (5e.2.7) を

$$\inf_{|\theta-\theta^0|=\delta} \sum p_i \log \frac{\pi_i(\theta^0)}{\pi_i(\theta)} > 0$$

と書きかえてもよいので，球面 $|\theta-\theta^0|=\delta$ 上のすべての点にたいして

$$\sum p_i \log \pi_i(\theta^0) > \sum p_i \log \pi_i(\theta)$$

である．それゆえ，開球 $|\theta-\theta^0|<\delta$ のなかに，関数 $\sum p_i \log \pi_i(\theta)$ が極大値をとる点 $\bar{\theta}$ が存在する．$\pi_i(\theta)$ は微分可能であるから，$\bar{\theta}$ での $\sum p_i \log \pi_i(\theta)$ の微係数は零になる．δ は任意であるから，$\bar{\theta}$ は尤度方程式の，一致性をもつ解である．

m.l. 方程式は

$$n\sum \frac{p_i}{\pi_i}\frac{\partial \pi_i}{\partial \theta_r}=0, \qquad r=1,\cdots,q \tag{5e.2.8}$$

である．$\sqrt{n}$ で割れば，r 番目の方程式は

$$\sum \frac{\sqrt{n}(p_i-\pi_i^0)}{\bar{\pi}_i}\frac{\partial \pi_i}{\partial \bar{\theta}_r}=\sum \frac{\sqrt{n}(\bar{\pi}_i-\pi_i^0)}{\bar{\pi}_i}\frac{\partial \pi_i}{\partial \bar{\theta}_r} \tag{5e. 2. 9}$$

と書くことができる．$\bar{\pi}_i-\pi_i^0$ を θ^0 で展開すれば，

$$\bar{\pi}_i-\pi_i^0=\sum(\bar{\theta}_s-\theta_s^0)\frac{\partial \pi_i}{\partial \theta_s'}, \qquad \theta_s'\in(\bar{\theta}_s,\ \theta_s^0)$$

を得るので，これを (5e. 2. 9) の右辺に代入すれば，

$$\sum_{i=1}^{k}\frac{\sqrt{n}(p_i-\pi_i^0)}{\bar{\pi}_i}\frac{\partial \pi_i}{\partial \bar{\theta}_r}=\sum_{s=1}^{q}\sqrt{n}(\bar{\theta}_s-\theta_s^0)j_{rs} \tag{5e. 2. 10}$$

であることがわかる．ここで

$$j_{rs}=\sum\frac{1}{\bar{\pi}_i}\frac{\partial \pi_i}{\partial \bar{\theta}_r}\frac{\partial \pi_i}{\partial \theta_s'}\longrightarrow i_{rs}, \qquad \bar{\theta}_r\to\theta_r^0 \text{ のとき}$$

である．$\sqrt{n}(p_i-\pi_i^0)$ は極限分布をもち，$\bar{\pi}_i\longrightarrow\pi_i^0$ かつ $(\partial\pi_i/\partial\bar{\theta}_r)\longrightarrow(\partial\pi_i/\partial\theta_r^0)$ であるので，結果 [(x)，**2c. 4**] を適用すれば，(5e. 2. 10) の左辺は漸近的に

$$\sum\frac{\sqrt{n}(p_i-\pi_i^0)}{\pi_i^0}\frac{\partial \pi_i}{\partial \theta_r^0}=\sqrt{n}z_r$$

と同等であることがわかる．したがって，関係式（確率的には漸近的に同等であることを記号 $\stackrel{a}{=}$ であらわす）

$$\sum_{s=1}^{q}\sqrt{n}(\bar{\theta}_s-\theta_s^0)j_{rs}\stackrel{a}{=}\sqrt{n}z_r \tag{5e. 2. 11}$$

が導かれる．関係式 (5e. 2. 11) を逆転すれば，$j^{rs}\longrightarrow i^{rs}$ であるので

$$\sqrt{n}(\bar{\theta}_s-\theta_s^0)\stackrel{a}{=}\sum\sqrt{n}z_rj^{rs}\stackrel{a}{=}\sum\sqrt{n}z_ri^{rs} \tag{5e. 2. 12}$$

が得られ，証明ができた．

注意 1.　(iv) の命題において，仮定 1.2 を仮定 1.1 におきかえれば，m.l. 推定量は尤度方程式の一致性をもつ解であるので，性質 (5e. 2. 5) および漸近分布について導かれた結果はそのまま m.l. 推定量について成立する．

注意 2.　Birch (1964) は，m.l. 推定量の漸近正規性を導くためには，$\pi_i(\theta)$ の1次導関数が連続であるという仮定 2.3 は，各 i にたいして $\pi_i(\theta)$ が全微分可能であるというより弱い仮定 2.4 におきかえられることを示した．より正確には，バーチの結果は次の通りである．仮定 1.1，2.4 および 3 が成り立てば，m.l. 推定量は (5e. 2. 5) の意味で有効であり，その漸近分布も (iv) で述べた通りである．証明は多少複雑であり詳細は Birch (1964) に与えられている．(iv) の結果は，m.l. 推定量であるか否かに関わらず，尤度方程式の根について述べているという意味で，より一般的であるが，より強い仮定 2.3 を用いている．

注意 3.　**5e. 1** 節および **5e. 2** 節の結果は，推定に関する他の文献に見られる仮定よりも，ずっと弱い仮定のもとで証明されている．母数の m.l. 推定量の漸近正規性が，θ の関数としての各群の出現確率の1次微分の連続性ないしは全微分可能性だけを仮定して導かれることは，注目すべき結果である．

有限の多項分布について示された結果を，可算個の値をとる一般の離散分布に拡張することは可能である．たとえば，$\pi=(\pi_1,\pi_2,\cdots)$ で許容集合 A に属する一般の離散分布をあらわし，π^0 が真の分布であるとすると，$\sum\pi_i^0\log\pi_i^0$ が有限値をとるという条件が成立すれば，近似 m.l. 推定量は一致性をもつことが保証される (Rao (1957a, 1958c), Kiefer and Wolfowitz (1956))．母数の推定のためには，それが π の連続な汎関数であるという仮定が必要である．この条件と π_i が θ の関数として連続であるという条件のもとでは，θ の m.l. 推定量が確率1で存在することが示される．π_i が θ の微分可能な関数であれば，m.l. 推定量はまた m.l. 方程式推定量でもある．しかしながら，推定量の漸近正規性が成り立つためには，さらに仮定が必要である (Rao (1958c))．

5f　一般の場合における母数の推定

5f.1　仮定と記号

連続型の場合の分布関数の推定については，母数を推定することによる方法を除いてはほとんどわかっていない．母数の推定による場合ですら，問題は複雑である．ここでは，母数が1つの場合だけを考え，それが成り立つために確率密度に関しあまり多くの仮定をおくことは必要としないような結果をいくつか証明する．

x を，σ 有限な測度 μ に関して確率密度 $p(\cdot,\theta)$ をもつ確率変数とする．ここで θ は実軸上の退化していない区間に属する．離散型の確率変数にたいしては，$p(y,\theta)$ は $x=y$ の確率をあらわす．いま，次の仮定をおく．

仮定 1.　導関数 $d\log p/d\theta$, $d^2\log p/d\theta^2$, $d^3\log p/d\theta^3$ が，真値を含む θ のある区間 A において，ほとんどすべての x にたいして存在する．

仮定 2.　θ の真値において

$$E\left[\frac{p'(x,\theta)}{p(x,\theta)}\Big|\theta\right]=0, \qquad E\left[\frac{p''(x,\theta)}{p(x,\theta)}\Big|\theta\right]=0,$$

$$E\left[\frac{p'(x,\theta)^2}{p(x,\theta)}\Big|\theta\right]>0$$

である．ここで $'$ の記号は θ についての微分をあらわす．

仮定 3.　A に属するすべての θ にたいして

$$\left|\frac{d^3\log p}{d\theta^3}\right| < M(x), \qquad E[M(x)|\theta] < K$$

である．ここでKはθに無関係である．

$x_1, \cdots, x_n$ を n 個の独立な観測値（確率変数）とし，

$$l(\theta|x) = \sum \log p(x_i, \theta), \qquad \frac{dl}{d\theta} = \sum \frac{p'(x_i, \theta)}{p(x_i, \theta)}$$

とする．

5f.2　m.l. 方程式推定量の性質

（i）　$\log p(x, \theta)$ が真値を含むある区間で微分可能ならば，m.l. 方程式は $n \to \infty$ のとき確率1でθの一致推定量である根をもつ．

θ_0 を真値とし，2つの値 $\theta_0 \pm \delta$ を考える．不等式（1e.6.6）を用いて

$$E\left\{\log \frac{p(\theta_0 - \delta)}{p(\theta_0)}\middle|\theta_0\right\} < 0, \qquad E\left\{\log \frac{p(\theta_0 + \delta)}{p(\theta_0)}\middle|\theta_0\right\} < 0 \tag{5f.2.1}$$

を得る．(5f.2.1) の厳密な不等号は，$p(x, \theta)$ が θ_0 を含むある区間でθに無関係でない限り成立させることができる．$p(x, \theta)$ が θ_0 を含むある区間で θ に無関係である場合には，証明しようとしている結果は明らかに成立する．

不等式（5f.2.1）より，大数の強法則を適用すれば，確率1で，$n \to \infty$ のとき

$$\frac{1}{n}[l(\theta_0 \pm \delta|x) - l(\theta_0|x)] < 0 \tag{5f.2.2}$$

である．つまり，ほとんどすべての標本列にたいして，$l(\theta|x)$ は $\theta_0 + \delta$ および $\theta_0 - \delta$ での値よりも θ_0 での値の方が最終的には大きいのである．$l(\theta|x)$ は $(\theta_0 \pm \delta)$ で連続であるから，$(\theta_0 \pm \delta)$ のなかに $l(\theta|x)$ の極大値が存在する．$l(\theta|x)$ が微分可能ならば，その点での微係数は零でなければならない．δは任意であり，根$\bar\theta$は $(\theta_0 \pm \delta)$ のなかにあるのだから，$\bar\theta$ は θ_0 の一致推定量である．

（ii）　$\bar\theta$ を尤度方程式の一致性をもつ根とし，仮定 1—3 が成立するものとする．そうすると，

$$\sqrt{n}\left[(\bar\theta - \theta_0)i(\theta_0) - \frac{1}{n}\frac{dl}{d\theta_0}\right] \to 0, \qquad \text{確率収束の意味で} \tag{5f.2.3}$$

が成立する．

結果（5f.2.3）は，$\bar\theta$ が（5c.2.5）の意味で f.o.e. であり，$\sqrt{n}(\bar\theta - \theta_0)$ の漸近分布が $N(0, i^{-1}(\theta_0))$ であることを示している．

$\bar\theta$ は尤度方程式 $(dl/d\bar\theta) = 0$ を満足するので，$(dl/d\bar\theta)$ を θ_0 でテイラー展開すれば，

$$0 = \frac{dl}{d\theta_0} + (\bar\theta - \theta_0)\frac{d^2l}{d\theta_0^2} + \frac{(\bar\theta - \theta_0)^2}{2}\frac{d^3l}{d\theta_1^3}, \qquad \theta_1 \in (\theta_0, \bar\theta) \tag{5f.2.4}$$

を得る．これを書きかえれば，方程式

$$\sqrt{n}(\bar\theta - \theta_0) = \frac{-\dfrac{1}{\sqrt{n}}\dfrac{dl}{d\theta_0}}{\dfrac{1}{n}\left\{\dfrac{d^2l}{d\theta_0^2} + \dfrac{\bar\theta - \theta_0}{2}\dfrac{d^3l}{d\theta_1^3}\right\}} \tag{5f.2.5}$$

が導かれる．ところで，

$$\sqrt{n}(\bar\theta - \theta_0)i(\theta_0) - \frac{1}{\sqrt{n}}\frac{dl}{d\theta_0} = -(b_n + 1)\frac{1}{\sqrt{n}}\frac{dl}{d\theta_0} \tag{5f.2.6}$$

を考える．ここで

$$b_n = \frac{i(\theta_0)}{\dfrac{1}{n}\left\{\dfrac{d^2l}{d\theta_0^2} + \dfrac{\bar\theta - \theta_0}{2}\dfrac{d^3l}{d\theta_1^3}\right\}}.$$

もし確率1で $b_n \to -1$ であることが示されれば，$(1/\sqrt{n})(dl/d\theta_0)$ がある漸近分布に収束することから，(5f.2.6) の右辺が零に確率収束するので，証明が完結する．

いま，仮定2のもとでは

$$E\left\{\frac{d^2\log p}{d\theta_0^2}\middle|\theta_0\right\} = -i(\theta_0)$$

であることがわかるので，大数の強法則を適用すれば

$$\frac{1}{n}\frac{d^2l}{d\theta_0^2} \to -i(\theta_0) \tag{5f.2.7}$$

が得られる．また，仮定3のもとでは，$n \to \infty$ のとき確率1で

$$\frac{1}{n}\left|\frac{d^3l}{d\theta_1^3}\right| < K$$

であり，$(\bar\theta - \theta_0) \to 0$ であるので，収束についての結果［(x), **2c.4**］を用いて

$$(\bar\theta - \theta_0)\frac{1}{n}\left|\frac{d^3l}{d\theta_1^3}\right| \to 0, \qquad \text{確率1で} \tag{5f.2.8}$$

である．(5f.2.7) と (5f.2.8) を合わせれば，確率1で $b_n \to -1$ である．

（iii）　$\bar\theta_1$ と $\bar\theta_2$ を尤度方程式の2つの一致性をもつ根とすると，それらは

$$\sqrt{n}(\bar\theta_1 - \bar\theta_2) \to 0, \qquad \text{確率収束の意味で} \tag{5f.2.9}$$

という意味で同等である．

$\bar\theta_1$ と $\bar\theta_2$ は双方とも有効推定量であるので，結果［(ii), **5c.2**］を，$\gamma(\theta) = 1$ であることに注意して適用すれば，この結果が導かれる［同等な結果が Huzurbazar (1948) に与えられているので参照せよ］．

この節の結果（i），(ii)，(iii) によっては，尤度方程式の一致性をもつ根を識別することはできない．一致性をもつ根が確率1で尤度の上限に対応することが示せれば，より満足できる結果であろう．これが成立するためには，密度関数が，Wald (1949) によって与えられたような条件をもみたさなければならない．m.l. 推定量の漸近的性質の厳密な取り扱いについては，Le Cam (1953, 1956) を参照せよ．大標本における推定についての，他

の興味深い論文として Bahadur (1958), Daniels (1961), Dugue (1937), Haldane (1951), Neyman (1949) がある.

5g　母数の推定における得点による方法

最尤方程式は通常複雑であるので，解を直ちに得ることができない．一般的な方法は，試験的な解を仮定し，小さな付加的修正量についての線形方程式を導くことである．この手順は，修正量が無視できるほどになるまで繰り返される．付加的修正量についての線形方程式を得るのに，得点による方式として知られる方法を採用すれば，これを機械的に行うことができる．

L を母数 θ の尤度としたとき，$d\log L/d\theta$ を θ についての有効得点と定義する．最尤推定値は，有効得点が零になるような θ の値である．θ_0 を推定値の試験的な値としたとき，$d\log L/d\theta$ を展開し $\delta\theta = \theta - \theta_0$ の1次までの項だけを残せば

$$\frac{d\log L}{d\theta} \simeq \frac{d\log L}{d\theta_0} + \delta\theta\frac{d^2\log L}{d\theta_0^2}$$

$$\simeq \frac{d\log L}{d\theta_0} - \delta\theta\mathscr{I}(\theta_0)$$

である．ここで $\mathscr{I}(\theta_0)$ は $\theta = \theta_0$ のときの情報であって，$-d^2\log L/d\theta_0^2$ の期待値である．大標本では $-\mathscr{I}(\theta_0)$ と $d^2\log L/d\theta_0^2$ との差は，n を観測値の数とすれば $0(1/n)$ であろう．よって上の近似は微小量の1次の項まで正確である．修正量 $\delta\theta$ は方程式

$$\delta\theta\mathscr{I}(\theta_0) = \frac{d\log L}{d\theta_0}, \qquad \delta\theta = \frac{d\log L}{d\theta_0} \div \mathscr{I}(\theta_0)$$

から求められる．第1段階の近似値は $(\theta_0 + \delta\theta)$ であり，これを新しい試験的な値として，この手順が繰り返される．

例 1.　コーシー分布

$$\frac{1}{\pi}\frac{dx}{1+(x-\theta)^2}$$

からの大きさ n の標本を考える．尤度方程式は

$$\frac{d\log L}{d\theta} = \sum\frac{2(x_i-\theta)}{1+(x_i-\theta)^2} = 0$$

である．任意の θ についての有効得点は

$$S(\theta) = \sum\frac{2(x_i-\theta)}{1+(x_i-\theta)^2}$$

である．ただ1つの観測値に含まれる情報は1/2であるので，$\mathscr{I}(\theta) = n/2$ であるから，

試験的な値 θ_0 に加える修正量は単に $2S(\theta_0)/n$ である．この手順が，安定した値に到達するまで，続けられる．この例では幸運にも $\mathscr{I}(\theta)$ が θ に無関係であるので，試験的な値各々について $\mathscr{I}(\theta)$ を計算する必要はない．

例 2.　**群分けしたデータについての得点と情報**　$\pi_1, \cdots, \pi_k$ $(\sum\pi_i = 1)$ を k 個の互いに排反なクラスの出現確率とし，ただ1つの母数 θ の関数として定義されているものとする．$f_1, \cdots, f_k$ $(\sum f_i = f)$ を観測度数とすれば，

$$\log L = f_1\log\pi_1 + \cdots + f_k\log\pi_k$$

である．θ での得点は

$$\frac{d\log L}{d\theta} = \frac{f_1}{\pi_1}\frac{d\pi_1}{d\theta} + \cdots + \frac{f_k}{\pi_k}\frac{d\pi_k}{d\theta}$$

$$\mathscr{I}(\theta) = V\left(\frac{d\log L}{d\theta}\right) = f\sum\frac{1}{\pi_i}\left(\frac{d\pi_i}{d\theta}\right)^2$$

である．ところで

$$\frac{1}{\pi_i}\frac{d\pi_i}{d\theta} \quad と \quad \frac{1}{\pi_i}\left(\frac{d\pi_i}{d\theta}\right)^2$$

とは，各々 i 番目のクラスの得点と情報と呼んでよい量であり，個々の問題では推定の手順を進める前に計算されるだろう．

遺伝学における連鎖の推定　2つの因子が組み替え率 θ で連鎖しているものとすれば，交配 $AB/ab \times AB/ab$ (相引的) と $Ab/aB \times Ab/aB$ (反発的) とにより，表 5g.1α に示したような期待比率と情報とが得られる．情報の量は，一方の交配の型が，他の型に比べて組み替え率の推定のために，どれほどの効率をもつかを判断するのに用いられる．

表 **5g.1α**

クラス	相引的な場合		反発的な場合	
	確率	得点	確率	得点
	4π	$\frac{1}{\pi}\frac{d\pi}{d\theta}$	4π	$\frac{1}{\pi}\frac{d\pi}{d\theta}$
AB	$3-2\theta+\theta^2$	$\frac{-2(1-\theta)}{3-2\theta+\theta^2}$	$2+\theta^2$	$\frac{2\theta}{2+\theta^2}$
Ab	$2\theta-\theta^2$	$\frac{2(1-\theta)}{2\theta-\theta^2}$	$1-\theta^2$	$\frac{-2\theta}{1-\theta^2}$
aB	$2\theta-\theta^2$	$\frac{2(1-\theta)}{2\theta-\theta^2}$	$1-\theta^2$	$\frac{-2\theta}{1-\theta^2}$
ab	$1-2\theta+\theta^2$	$\frac{-2(1-\theta)}{1-2\theta+\theta^2}$	θ^2	$\frac{2}{\theta}$
情報	$\frac{2(3-4\theta+2\theta^2)}{\theta(2-\theta)(3-2\theta+\theta^2)}$		$\frac{2(1+2\theta^2)}{(2+\theta^2)(1-\theta^2)}$	

たとえば，$\theta = 1/4$ のときには，相引的な場合と反発的な場合の情報の量は各々 3. 7909 と 1. 1636 である．このことは，同じ精度で組み替え率を推定するには，反発的な交配を用いた場合に必要な子孫の数は，相引的な場合の数の3倍になることを意味する．

下に示した通りの AB, Ab, aB, ab についての値をもつ，相引的な場合の資料を考える．$\theta = 0.21$ を試験的な値とすれば，次のようにして得点や情報が得られる．

	4π	$4\dfrac{d\pi}{d\theta}$	$\dfrac{1}{\pi}\dfrac{d\pi}{d\theta}$	観測度数
AB	2. 6241	−1. 58	−0. 60211	125
Ab	0. 3759	1. 58	4. 20325	18
aB	0. 3759	1. 58	4. 20325	20
ab	0. 6241	−1. 58	−2. 53164	34
	絶対値の和		11. 54025	197

1つの観測値あたりの情報：

$$\sum \frac{d\pi_i}{d\theta}\left(\frac{1}{\pi_i}\frac{d\pi_i}{d\theta}\right) = \frac{1.58}{4}(11.54025) = 4.55840$$

$$\text{有効得点} = -0.60211(125) + 4.20325(18+20) - 2.53164(34)$$
$$= -1.61601$$

$$\text{修正項} = -\frac{1.61601}{197(4.55840)} = -0.0018$$

$$\text{第1段階の近似値} = 0.21 - 0.0018 = 0.2082.$$

修正量は小さいので，手順を繰り返さなくてもよいだろう．推定値の精度を決める分散は $1/197(4.55840) = 0.00111357$ で与えられる．

分散のよりよい推定値は，$\theta = 0.2082$ のときの情報の逆数である．

1％から50％までの試験的な値についての得点が F. Y. 数表に与えられている．その数表は，近似の各段階で小数点以下2桁の数値が得られれば使用できるが，2桁の数値が安定してしまったときには，もっと桁をとって，数値例で示したように，完全な計算の手順を踏まなければならない．

例 3.　複数個の母数についての得点　例2で示した得点による方法は，複数個の母数の同時推定の場合に拡張できる．$\theta_1, \theta_2, \cdots, \theta_q$ を母数とすれば，i 番目の有効得点は

$$S_i = \frac{\partial \log L}{\partial \theta_i}, \qquad i = 1, 2, \cdots, q$$

により定義される．ここで L は母数の尤度である．そして情報行列は $(\mathscr{I}_{ij})$ により定義される．ここで

$$\mathscr{I}_{ij} = E(S_i S_j).$$

試験的な値 $\theta_1^0, \cdots, \theta_q^0$ についての有効得点および情報の値を添字零をつけてあらわせば，試験的な値にたいする小さな付加的修正量 $\delta\theta_1, \cdots, \delta\theta_q$ は，連立方程式

$$\mathscr{I}_{11}^0\delta\theta_1 + \mathscr{I}_{12}^0\delta\theta_2 + \cdots + \mathscr{I}_{1q}^0\delta\theta_q = S_1^0$$
$$\cdots \qquad \cdots \qquad \cdots \qquad \cdots$$
$$\mathscr{I}_{q1}^0\delta\theta_1 + \mathscr{I}_{q2}^0\delta\theta_2 + \cdots + \mathscr{I}_{qq}^0\delta\theta_q = S_q^0$$

の解として与えられる．この手順が，$\theta_1, \cdots, \theta_q$ の安定値が得られるまで，各回毎に修正された値を用いて繰り返される．θ_i の最終的推定値 $\hat\theta_i$ の分散は，$\mathscr{I}^{ii}$，つまり $(\mathscr{I}_{ij})$ の逆行列の i 番目の対角要素によって与えられる．

この方法で最もめんどうな点は，近似の各段階で情報行列を計算し，その逆行列を求めることである．実際には，このことは不必要であることがわかる．情報行列は，数段階経た後には固定しておき，得点だけを再計算する．この手順によれば，かなり計算量を減らせる．安定した値に到達した最終段階で，推定値の分散と共分散を得るために，その推定値を用いて情報行列を計算すればよい．

π_i と f_i をそれぞれ i 番目のクラスの出現確率と観測度数とする多項分布では，

$$S_t = \sum_i \frac{f_i}{\pi_i}\frac{\partial \pi_i}{\partial \theta_t}$$

であり，

$$\mathscr{I}_{ut} = f\sum_i \frac{1}{\pi_i}\frac{\partial \pi_i}{\partial \theta_u}\frac{\partial \pi_i}{\partial \theta_t}$$

であるので，次の節で示す通り，計算は簡単になる．

血液型の $A\ B\ O$ システムにおける遺伝子の比率の推定　すべての人間は，4つの血液型のグループ O，A，B，AB のどれかに分類される．これらの血液型の遺伝は，3つの対立（形質の）遺伝子 O，A，B によって決まるが，そのうち O は A，B にたいして劣性である．r, p, q を各々遺伝子 O, A, B の比率とすれば，無作為に交配させたときの，6つの遺伝子型（4つの表現型）の確率は次の通りである．

表現型	遺伝子型	確率
O	OO	r^2
A	$\left.\begin{matrix}AA\\AO\end{matrix}\right\}$	$\left.\begin{matrix}p^2\\2pr\end{matrix}\right\}p^2+2pr$
B	$\left.\begin{matrix}BB\\BO\end{matrix}\right\}$	$\left.\begin{matrix}q^2\\2qr\end{matrix}\right\}q^2+2qr$
AB	AB	$2pq$

$\overline{O}, \overline{A}, \overline{B}, \overline{AB}$ を観測度数とし，合計が N であるとしたとき，問題は遺伝子の比率 p，q，

r を推定することである．1つの大まかな推定値は，

$$r' = \sqrt{\frac{\bar{O}}{N}}$$

$$p' = 1 - \sqrt{\frac{\bar{O} + \bar{B}}{N}}$$

$$q' = 1 - \sqrt{\frac{\bar{O} + \bar{A}}{N}}$$

で与えられる．これらは加えても必ずしも1にはならないが，真値はそうでなければならない．D でその差をあらわす．つまり

$$-D = p' + q' + r' - 1.$$

バーンシュタインによるよりよい推定値が次のようにして得られる．

$$r = \left(1 + \frac{1}{2}D\right)\left(r' + \frac{1}{2}D\right)$$

$$p = \left(1 + \frac{1}{2}D\right)p', \qquad q = \left(1 + \frac{1}{2}D\right)q'.$$

これでも，依然として差が $1 - p - q - r = \frac{1}{4}D^2$ だけ残る．この量が小さければ，バーンシュタインの方法はかなりよい推定値を与える．いま，これらの推定値が最尤法によってどのように改良されるかを，度数が $\bar{O} = 176$, $\bar{A} = 182$, $\bar{B} = 60$, $\overline{AB} = 17$ の場合について示そう．バーンシュタインの方法により得られる近似解は

$$p_0 = 0.26449, \qquad q_0 = 0.09317, \qquad r_0 = 0.64234$$

である．一般に，確率と，独立な母数 p と q についてのその導関数は次の通りである．

	確　率	導関数	
	π	$\frac{\partial\pi}{\partial p}$	$\frac{\partial\pi}{\partial q}$
O	r^2	$-2r$	$-2r$
A	$p(p+2r)$	$2r$	$-2p$
B	$q(q+2r)$	$-2q$	$2r$
AB	$2pq$	$2q$	$2p$

確率と，上に得られた近似値について有効得点を計算するための係数を表 5g. 1β に示す．

表 5g. 1β　得点計算のための係数

	確　率 π	$\frac{1}{\pi}\frac{\partial\pi}{\partial p}$	$\frac{1}{\pi}\frac{\partial\pi}{\partial q}$	観測度数
O	0. 41260	−3. 11362	−3. 11362	176
A	0. 40974	3. 13543	−1. 29104	182
B	0. 12838	−1. 45217	10. 00685	60
AB	0. 04928	3. 78086	10. 73307	17
				$435 = N$

得点は

$$\phi_p = (-3.11362)176 + (3.13543)182 + (-1.45217)60 + (3.78086)17 = -0.20444$$

$$\phi_q = (-3.11362)176 + (-1.29104)182 + (10.00685)60 + (10.73307)17 = -0.09324$$

である．1つの観測値についての情報行列は

$$i_{pp} = \sum\left(\frac{\partial\pi_i}{\partial p}\right)\left(\frac{1}{\pi_i}\frac{\partial\pi_i}{\partial p}\right) = 9.00315, \qquad i_{pq} = \sum\left(\frac{\partial\pi_i}{\partial p}\right)\left(\frac{1}{\pi_i}\frac{\partial\pi_i}{\partial q}\right) = 2.47676$$

$$i_{pq} = 2.47676, \qquad i_{qq} = \sum\left(\frac{\partial\pi_i}{\partial q}\right)\left(\frac{1}{\pi_i}\frac{\partial\pi_i}{\partial q}\right) = 23.2162$$

である．1つの観測値あたりの情報行列の逆行列は

$$i^{pp} = 0.114430 \qquad i^{pq} = -0.012208$$

$$i^{pq} = -0.012208 \qquad i^{qq} = 0.044376$$

である．修正量についての解は

$$\delta p = \frac{i^{pp}\phi_p + i^{pq}\phi_q}{N} = -0.00005116$$

$$\delta q = \frac{i^{pq}\phi_p + i^{qq}\phi_q}{N} = -0.00000377$$

である．この場合には修正はほとんど不必要である．修正量が小さくないならば，第2の近似値を用いて全手順を繰り返さなければならない．何段階か経た後には，各近似にたいし情報行列を再計算する必要がないことに注意することは大切である．新しい得点だけを各段階で計算し，より精密な近似を得るために，情報行列の逆行列（ある段階以後は一定のままである）と共に，用いなければならない．収束が確認されたときには，最終的近似値を得るために，情報行列と得点を計算する．

最尤推定値と分散は

$$\hat{p} = 0.26444, \qquad V(\hat{p}) = \frac{i^{pp}}{N} = 0.00026305$$

$$\hat{q} = 0.09317, \qquad V(\hat{q}) = \frac{i^{qq}}{N} = 0.00010202$$

$$\hat{r} = 0.64239, \qquad V(\hat{r}) = \frac{i^{pp} + 2i^{pq} + i^{qq}}{N} = 0.00030893$$

である．

複雑な問題での，得点による方法のこれ以外の応用例については，Finney (1952), Fisher (195[illegible]), Rao (1950a) を参照されたい．

O, A, B 血液型のデータの場合には，m, l. 推定値を得る別のアルゴリズムがあり，この方法の方が卓上あるいは電子計算機には適しているようであり，情報行列とその逆行列を

繰り返し計算することも避けられる．

p_0, q_0, r_0 を仮の推定値とし，$P_0 = p_0/r_0$ と $Q_0 = q_0/r_0$ とを計算する．p/r と q/r の k 番目の近似値を P_k と Q_k であらわすことにする．$k+1$ 番目の近似値は公式

$$\frac{\bar{A}+\overline{AB}}{P_{k+1}} = 2(\bar{O}) + \frac{\bar{A}}{2+P_k} + \frac{2\bar{B}}{2+Q_k}$$

$$\frac{\bar{B}+\overline{AB}}{Q_{k+1}} = 2(\bar{O}) + \frac{2\bar{A}}{2+P_{k+1}} + \frac{\bar{B}}{2+Q_k}$$

から計算される．P と Q の安定した値が得られるまで反復を行い，その安定値 P, Q を用いて，p, q, r の推定値は

$$\hat{r} = \frac{1}{P+Q+1}, \qquad \hat{p} = \hat{r}P, \qquad \hat{q} = \hat{r}Q$$

として計算される．仮の推定値として

$$p_0 = 0.26449, \qquad q_0 = 0.09317, \qquad r_0 = 0.64234$$

をとり $P_0 = 0.41176$, $Q_0 = 0.14505$ とすれば，最初の一巡の計算により $P_1 = 0.41167$, $Q_1 = 0.14503$ となり，安定した値が得られたと考えられる．p, q, r の m.l. 推定値は

$$\hat{r} = 0.64238, \qquad \hat{p} = 0.26445, \qquad \hat{q} = 0.09316$$

であり，情報行列を用いて反復手順により得られた推定値とほとんど一致している．こうして得られた推定値を用いて，各血液型の期待比率や情報行列が，推定値の標準誤差を計算するために前と同様にして求められる．

データの結合　得点による方法の利点は，同じ母数についての情報を与えるデータの集まりがいくつかあるとき，推定のためにそれらを結合することが機械的にできることにある．L をすべてのデータに基づく同時尤度とし，L_i を i 番目の部分についての尤度とすれば，

$$L = L_1 L_2 \cdots$$

$$\frac{\partial \log L}{\partial \theta_r} = \frac{\partial \log L_1}{\partial \theta_r} + \frac{\partial \log L_2}{\partial \theta_r} + \cdots$$

であるので，有効得点が加法的であることがわかる．また，$\mathscr{I}_{rs}$ を全データについての情報行列の要素とし，$\mathscr{I}_{rs}^{(j)}$ を j 番目の部分についてのそれとすれば，

$$\mathscr{I}_{rs} = \mathscr{I}_{rs}^{(1)} + \mathscr{I}_{rs}^{(2)} + \cdots$$

である．したがって，最良の推定値を得るためには，データの各部分を，試験的な値での得点と情報行列に変換し，全体での得点と情報行列を単なる足し算で求めることが必要である．試験的な値にたいする修正量は，数値例で示したように，同時方程式を解くことにより得ることができる．

補足と問題

1　範囲が0から θ までの長方形分布を考える．$x_{\max}$ を，この分布からの n 個の独立な観測値の最大値とする．そのとき，次のことを示せ．

1.1　$x_{\max}$ は θ の完備十分統計量である．

1.2　$(n+1)x_{\max}/n$ は θ の m.v.u.（最小分散不偏）推定量である．最小分散の値も求めよ．

1.3　$x_{\max}$ は θ の m.l. 推定量であり，θ の不偏推定量ではないが一致推定量である．

1.4　θ の不偏推定量である，$x_{\max}$ の任意の関数 $t(x_{\max})$ は，$t(\theta) = (n+1)\theta/n$ という値をとる．

2　$x_{\max}$ と $x_{\min}$ を，範囲が θ から ϕ までの長方形分布からの n 個の観測値の最大値と最小値とする．そのとき，次のことを示せ．

2.1　$x_{\max}$ と $x_{\min}$ は θ と ϕ の完備十分統計量である．

2.2　中央点 $(\theta+\phi)/2$ と範囲 $\phi-\theta$ の m.v.u. 推定量は各々

$$\frac{x_{\max}+x_{\min}}{2} \quad と \quad \frac{n+1}{n-1}(x_{\max}-x_{\min})$$

である．

2.3　$\phi = 2\theta$ とすると母数は1つになる．しかし，依然として $x_{\max}$ と $x_{\min}$ とが θ の十分統計量であるが，完備ではない．

2.4　$\phi = 2\theta$ のとき，θ の m.l. 推定量は $x_{\max}/2$ であるが，θ の不偏推定量ではない．m.l. 推定量を基にして得られる1つの不偏推定量，つまり，偏りを修正した推定量は，$T_1 = (n+1)x_{\max}/(2n+1)$ である．

2.5　$T_2 = (n+1)(2x_{\max}+x_{\min})/(5n+4)$ は，分散が T_1 のそれよりも一様に小さい θ の推定量であることを示せ．したがって，最小分散の規準により判断したときに，m.l. 推定量が劣っているという例を得る．

3　離散分布の確率の最小分散不偏推定量　整数値 $0, 1, \cdots$ をとる確率変数 X を考え，$P(X=r) = \pi_r(\theta)$ とする．ここで θ は未知母数である．$x_1, \cdots, x_n$ を X についての n 個の独立な観測値とし，T を θ の完備十分統計量とする．$\pi_r(\theta)$ を推定するのに，p =（その値が r に等しい x_i の個数）$/n$ という統計量を考える．

3.1　結果［(iii), 5a.2］を用いて，$\pi_r(\theta)$ の m.v.u. 推定量は $E(p|T)$ であることを示せ．

3.2　上の結果を適用して，n 個の独立な観測値 $x_1, \cdots, x_n$ に基づくポアソン確率 $P(X=0) = e^{-\mu}$ の m.v.u. 推定量を求めよ．

3.3　X が負の2項分布をするとき，n 個の独立な観測値に基づく $P(X=0)$ の m.v.u. 推定量を求めよ．

4　不偏推定量の分散についてのより厳密な不等式　T を $g(\theta)$ の不偏推定量とし，$P(X,\theta)$ を X での確率密度とする．関係式

$$\int TP(X,\theta)dv = g(\theta)$$

を k 回微分すれば，

$$\int T\frac{d^k P}{d\theta^k}dv = \frac{d^k g}{d\theta^k}$$

を得る．

4.1　$V(T) \geqslant [d^k g/d\theta^k]^2/J_{kk}$ であることを示せ．ここで

$$J_{kk} = \int \left(\frac{d^k P}{d\theta^k}\right)^2 \frac{1}{P} dv.$$

4.2　より一般には，$J_{kr} = \mathrm{cov}\left(\frac{1}{P}\frac{d^k P}{d\theta^k}, \frac{1}{P}\frac{d^r P}{d\theta^r}\right)$ とすれば，T と

$$\frac{1}{P}\frac{d^i P}{d\theta^i}, \qquad i = 1, 2, \cdots$$

との間の重相関係数は

$$\left[\sum\sum J^{kr}\frac{d^k g}{d\theta^k}\frac{d^r g}{d\theta^r}\right] \div V(T)$$

であることを示せ．ここで (J^{kr}) は (J_{kr}) の逆行列である．それにより，不等式

$$V(T) \geqslant \sum\sum J^{kr}\frac{d^k g}{d\theta^k}\frac{d^r g}{d\theta^r}$$

を導け．

4.3　問題 **4.2** で得られた下界は $[g'(\theta)]^2/\mathscr{I}(\theta)$ より小さくはないことを示せ．上の不等式は Bhattacharya (1946) によるものである．

5　双生児には一卵性かそうでないかの2つの型がある．π を子供が男子である確率とし，α を双生児が一卵性である確率とする．一卵性双生児は受精卵の分割により生まれるとすれば，双生児が型 ♂♂, ♂♀, ♀♀ である確率は各々，$\pi(\pi+\alpha\phi)$, $2\pi\phi(1-\alpha)$, $\phi(\alpha\pi+\phi)$ であることを示せ．ここで $\phi = 1-\pi$ である．n_1, n_2, n_3 を，合計 $n = n_1 + n_2 + n_3$ 組の双生児からなる標本における，これらの型の観測度数とする．

5.1　α, π の推定量を求めよ．

5.2　$\pi = 0.516$ のとき，α の推定量を m. l. 法および最小カイ2乗法により求めよ．

5.3　問題 **5.2** の推定量の標準誤差と，ウェインバーグの推定量 $\alpha^* = (1-n_2)/2n\pi\phi$ のそれとを比較せよ．

6　**ベキ級数分布の母数の m. v. u. 推定**　　確率が

$$P(X=r) = \frac{a_r}{f(\theta)}\theta^r, \qquad r = c, c+1, \cdots \infty$$

で与えられるベキ級数分布を考える．$x_1, \cdots, x_n$ を大きさ n の標本とし，総和を $T = x_1 + \cdots + x_n$ とする．

6.1　T は θ の完備十分統計量であり，T の分布は，X のそれと同様の形

$$P(T=t) = b_t\theta^t \div [f(\theta)]^n, \qquad t = nc, nc+1, \cdots \infty$$

であることを示せ．

6.2　θ^r の m. v. u. 推定量は

$$\begin{aligned} u_r(t) &= 0, & t &< r \\ &= \frac{b_{t-r}}{b_t}, & t &\geqslant r \end{aligned}$$

であり，その分散の m. v. u. 推定量は

$$[u_r(t)]^2 - u_{2r}(t)$$

であることを示せ [Roy and Mitra (1957)]．

[注意：問題 **6.1** および **6.2** は，左で切られたものを含めて範囲が無限区間にわたる，よく知られた離散分布をすべて包含する．]

6.3　問題 **6.1** および **6.2** の結果は，変量のとる値の範囲が無限でない場合には，成立しないことを示せ [Patil (1963)]．

7　**離散分布の母数の最尤推定量**　　ベキ級数分布

$$P(X=r) = a_r\theta^r/f(\theta), \qquad r = 0, 1, \cdots$$

を考える．ここで a_r はある r の値については零であってもよい．$x_1, \cdots, x_n$ を n 個の独立な観測値とし，$\bar{x} = (x_1 + \cdots + x_n)/n$ とする．

7.1　θ の m. l. 推定量は，方程式 $\bar{x} = \theta f'(\theta)/f(\theta) = \mu(\theta)$ の根であることを示せ．ここで $\mu(\theta)$ は理論上の平均である．推定方程式は，モーメント法についても同じである．

7.2　1つの観測値に含まれる情報 $i(\theta)$ は，$(d\mu/\theta d\theta) = \mu_2(\theta)/\theta^2$ であることを示せ．ここで $\mu_2(\theta)$ は分布の2次モーメントである．

7.3　ポアソン，2項分布，対数関数をベキ級数に展開することにより得られる分布，および零を除いたポアソンおよび2項分布について，母数を推定するための方程式を求めよ．

8　$x_1, \cdots, x_n$ を，確率密度 $p(x, \theta)$ をもつ確率変数についての n 個の独立な観測値とする．さらに，$f(x_1, \cdots, x_n, \theta) = 0$ を，θ についての推定方程式とする．$E(f) \equiv 0$ であれば，(積分記号のなかで微分し，1つの項に C-S 不等式を適用することにより)

$$V(f) \geqslant [E(f')]^2/\mathscr{I} \quad \text{あるいは} \quad [E(f')]^2/V(f) \leqslant \mathscr{I}$$

であることを示せ．ここで $f' = df(x_1, \cdots, x_n, \theta)/d\theta$ であり，$\mathscr{I}$ は標本に含まれる全情報である．$f = d\log L/d\theta$ と選べば，$[E(f')]^2/V(f)$ の上界に到達することを確かめよ．ここで $L = p(x_1, \theta) \cdots p(x_n, \theta)$ である．

[注意：この問題は，推定方程式 $f = 0$ の効率を $[E(f')]^2/V(f)$ で定義すれば，最尤推定方程式が最大の効率をもつことを示している．小標本における最尤法のよさを示そうとする試みが Godambe (1960) によりなされている]．

9　$x_1, \cdots, x_n$ を，分布関数 $F(x, \theta), \theta \in A$, をもつ確率変数についての n 個の独立な観測値からなる標本とする．ここで A は実軸上の退化していない区間とする．S_n で経験分布関数をあらわすものとする．そのとき，次のような型の推定方程式を考える．

$$\int g(x, \theta) dS_n = 0 = [g(x_1, \theta) + \cdots + g(x_n, \theta)] \div n,$$

ここで $g(x, \theta)$ は次の条件を満足するものとする．

(a)　$\int g(x, \theta) dF(x, \theta) \equiv 0,$

(b)　$\int \frac{dg(x, \theta)}{d\theta} dF(x, \theta) = m(\theta),$

(c)　$\int [g(x, \theta)]^2 dF(x, \theta) = h(\theta), \qquad 0 < h(\theta) < \infty,$

(d)　$\left|\frac{d^2 g(x, \theta)}{d\theta^2}\right| < M(x), \qquad E[M(x)|\theta] < K,$

ここで K は θ に無関係である．

9.1　方程式 $\sum g(x_i, \theta) = 0$ に，一致性をもつ根 θ^* が存在することを示せ．

9.2　**5f.2** の議論に従って，$\sqrt{n}(\theta^* - \theta)$ の漸近分布が $N(0, \sigma^2)$ であることを証明せよ．ここで $\sigma^2 = h(\theta)/[m(\theta)]^2$ である．

9.3　$F(x, \theta)$ が密度関数 $p(x, \theta)$ をもつものとする．適当な条件のもとで

$$\frac{h(\theta)}{[m(\theta)]^2} \geqslant \frac{1}{i(\theta)}$$

であることを示せ．ここで $i(\theta)=E[p'(x,\theta)/p(x,\theta)]^2$ である．

9.4　$g(x,\theta)=p'(x,\theta)/p(x,\theta)$ のとき，つまり，推定方程式が m.l. 方程式のとき，漸近分散の下界に到達することを検証せよ．

[したがって，ある推定方程式の集まりのなかでは，m.l. 法が最小の漸近分散をもつ推定量を与える方法であるとの特徴づけが得られる．]

10　**最尤法による分布の特徴づけ**

10.1　$\{F(x-\theta),\theta\in R^1\}$ を，移行性をもつ，実軸上の絶対連続な分布関数の族とし，その確率密度関数 $f(x)$ が $x=0$ で下方半連続であるものとする．大きさ2および3のすべての標本について，θ の1つの m.l. 推定値が標本の算術平均であれば，$F(x)$ は $N(0,\sigma^2)$ である．

10.2　$F(x/\sigma)$，$\sigma>0$ が絶対連続な尺度母数 σ をもつ分布の族を構成し，確率密度が次の条件をみたすものとする．

（a）　$f(x)$ は $(0,\infty)$ で連続である．

（b）　$y\downarrow 0$ のとき $\lim\dfrac{f(\lambda y)}{f(y)}=1$.

すべての標本にたいして σ の1つの m.l. 推定値が算術平均であれば，その分布関数は指数分布の族に属する．つまり

$$f(x)=e^{-x},\qquad x>0$$
$$f(x)=0,\qquad x\leqslant 0.$$

10.3　問題**10.2**において，$f(x)$ が条件

（a）　$f(x)$ は $(-\infty,\infty)$ で連続である．

（b）　$y\to 0$ のとき $\lim f(\lambda y)/f(y)=1$

を満足し，σ^2 の1つの最尤推定値が $\sum x_i^2/n$ であれば，その分布は $N(0,\sigma^2)$ である．

[注意：これらの命題は文献でははっきりとしないまま知られてはいたが，Teicher(1961)により厳密な形で述べられた．]

10.4　$\{F(x-\theta),\theta\in R^1\}$ を問題**10.1**で定義されたのと同じ族とする．$n=4$ のとき標本中央値が θ の m.l. 推定値であれば，$F'(x)=f(x)=(a/2)\exp(-a|x|)$，つまりラプラス分布である (Kagan, Linnik and Rao (1972B4) と Rao and Ghosh(1971)gを参照せよ)．

11　X を離散型の確率変数とし，$P(X=-1)=\alpha$, $P(X=n)=(1-\alpha)^2\alpha^n$, $n=0,1,\cdots$ とする．ここで $0<\alpha<1$ である．

11.1　X は α の最小十分統計量であるが，完備ではない $[E(X)=0]$ ことを示せ．

11.2　零の不偏推定量であるような X の関数は必ず cX という形でなければならない．それゆえ，$X=0$ のとき $T(X)=1$ で，その他のとき $T(X)=0$ という関数 $T(X)$ がすべての α にたいして $(1-\alpha)^2$ の m.v.u. 推定量であることを示せ．

11.3　α の不偏推定量は存在するが，一様 m.v.u. 推定量は存在しないことを示せ．

12　θ は2つの値0,1だけしかとらないものとし，確率密度を

$$p(x|\theta=0)=\begin{cases}1, & 0<x<1\\ 0, & \text{その他}\end{cases}$$

$$p^*(x|\theta=1)=\begin{cases}\frac{1}{2}x^{-\frac{1}{2}}, & 0<x<1\\ 0, & \text{その他}\end{cases}$$

で定義する．θ の不偏推定量は存在し，$\theta=0$ での分散はいくらでも小さくできるが，下限である零はどんな不偏推定量でも到達できないことを示せ．

参　考　文　献

Arnold, J. C. and S. K. Katti, (1972), An application of Rao-Blackwell theorem in preliminary test estimators, *J. Multivariate Analysis* **2**, 236–238.

Bahadur, R. R. (1958), Examples of inconsistency of maximum likelihood estimates, *Sankhyā* **20**, 207–210.

Bhattacharya, A. (1946), On some analogues of the amount of information and their uses in statistical estimation, *Sankhyā* **8**, 1–14, 201–218, 315–328.

Birch, M. W. (1964), A new proof of the Fisher-Pearson theorem, *Ann. Math. Statist.* **35**, 817–824.

Blackwell, D. (1947), Conditional expectation and unbiased sequential estimation, *Ann. Math. Statist.* **18**, 105–110.

Chapman Douglas, G. and Herbert Robbins (1951), Minimum variance estimation without regularity assumptions, *Ann. Math. Statist.* **22**, 581–586.

Cramer, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J.

Daniels, H. E. (1961), The asymptotic efficiency of a maximum likelihood estimator, *Proc. (Fourth) Berkeley Symp. on Math. Statist. Prob.* **1**, 151–163.

Dempster, A. P. (1963), On the difficulties inherent in Fisher's fiducial argument, *Proc. Inter. Stat. Conference*, Ottawa.

Doob, J. (1934), Probability and statistics, *Trans. Am. Math. Soc.* **36**, 759–775.

Doob, J. (1936), Statistical estimation, *Trans. Am. Math. Soc.* **39**, 410–421.

Dugue, D. (1937), Application des proprietes de la limite au sens du calcul des probabilities a l'etude des diverses questions d'estimation, *Ecol. Poly.* **3**, 305–372.

Finney, D. J. (1952), *Probit Analysis* (second edition), Cambridge University Press, Cambridge, Mass.

Fisher, R. A. (1922), On the mathematical foundations of theoretical statistics, *Philos. Trans. Roy. Soc.* **A222**, 309–368.

Fisher, R. A. (1925), Theory of Statistical estimation, *Proc. Camb. Phil. Soc.* **22**, 700–725.

Fisher, R. A. (1950), *Statistical Methods for Research Workers* (first edition 1925, eleventh edition 1950), Oliver and Boyd, Edinburgh.

Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd. Edinburgh and London.

Fraser, D. A. S. (1961), The fiducial method and invariance, *Biometrika* **48**, 261–280.

Gart, John J. (1959), An extension of the Cramer-Rao inequality, *Ann. Math. Statist.* **30**, 367–380.

Godambe, V. P. (1960), An optimum property of regular maximum likelihood estimation, *Ann. Math. Statist.* **31**, 1208–1212.

Godambe, V. P. and D. A. Sprott (1971), *Foundations of Statistical Inference*, Holt, Reinhart, and Winston, Canada.

Haldane, J. B. S. (1951), A class of efficient estimates of a parameter, *Bull. Int. Statist. Inst.* **33**, 231–248.

Hodges, J. L. (Jr.) and E. L. Lehmann (1951), Some applications of the Cramer-Rao inequality, *Proc. (Second) Berkeley Symp. on Math. Statist. Prob.*, 13–22.

Huzurbazar, V. S. (1948), The likelihood equation, consistency, and maxima of the likelihood function, *Ann. Eugen. (London)* **14**, 185-200.

Jeffreys, H. (1939), *Theory of Probability*, Clarendon Press, Oxford.

Kiefer, J. and J. Wolfowitz (1956), Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Ann. Math. Statist.* **27**, 887-906.

Kolmogorov, A. N. (1950), Unbiased estimates (in Russian), *Izvestia Acad. Nauk USSR* **14**, 303.

Le Cam, L. (1953), On some asymptotic properties of maximum likelihood estimates and related Bayes's estimates, *Uni. California Publ. Statist.* **1**, 277-330.

Le Cam, L. (1956), On the asymptotic theory of estimation and testing hypotheses, *Proc. (Third) Berkeley Symp. on Math. Statist. Prob.* **1**, 129-156.

Lehmann, E. L. and H. Scheffe (1950), Completeness, similar regions and unbiased estimation, *Sankhyā* **10**, 305-340.

Neyman, J. (1949), Contribution to the theory of the χ^2 test, *Proc. (First) Berkeley Symp. on Math. Statist. Prob.*, 239-273.

Patil, G. P. (1963), Minimum variance unbiased estimation and certain problems of additive number theory, *Ann. Math. Statist.* **34**, 1050-1056.

Perlman, M. D. (1972), Reduced mean square error estimation for several parameters, *Sankhyā*.

Robbins, H. (1955), An empirical Bayes approach to statistics, *Proc. (Third) Berkeley Symp. on Math. Statist. Prob.* **1**, 157-163.

Roy, Jogabratha and Sujit Kumar Mitra (1957), Unbiased minimum variance estimation in a class of discrete distributions, *Sankhyā* **18**, 317-378.

Sprott, D. A. (1961), Similarities between likelihoods and associated distributions *a posteriori*, *J. Roy. Statist. Soc.* **24**(B), 460-468.

Stein, C. (1950), Unbiased estimates with minimum variance, *Ann. Math. Statist.* **21**, 406-415.

Stein, C. M. (1962), Confidence sets for the mean of a multivariate normal distribution, *J. Roy. Statist. Soc.* **B24**, 265-296.

Teicher, H. (1961), Maximum likelihood characterization of distributions, *Ann. Math. Statist.* **32**, 1214-1222.

Wald, A. (1949), Note on the consistency of maximum likelihood estimate, *Ann. Math. Statist.* **20**, 595-601.

第6章　大標本の理論と方法

まえがき　この章では大標本が入手できる場合の統計的推測を扱う．その内容は，実際の応用のときに役立つ多種多様な統計的方法を含んでいる．特に重点をおいたのは，分類値（カテゴリカル・データ）の解析で，規定の問題，標本の一様性，属性の独立性などに関する問題である．大標本の場合の単純仮説や複合仮説のいくつかの一般的な検定基準についても論じた．また，検定の漸近効率については第7章の **7a.7** で述べる．

この章での理論的展開は，第2章で論じた基本的収束定理と第5章で扱った最尤推定量の漸近的諸性質に全面的に依拠している．読者は，大数の法則，1変量および多変量の中心極限定理，確率変数の関数の収束定理を取り扱った第2章の **2c.3** から **2c.5** の諸命題，および，推定量の漸近効率，大標本の場合の推定の方法，最尤法を扱った第5章の **5c** から **5f** の各節を学習されるとよい．

6a　基本的な諸結果

6a.1　度数の2次関数の漸近分布

k 個の互いに排反な事象があり，第 i 事象の生起確率を π_i $(i=1,\cdots,k)$ とする．いま n 個の独立な事象が観測され，第1事象の個数が n_1 個，第2事象が n_2 個，等々であり，$\sum n_i = n$ とする．列ベクトル $\mathbf{V}, \boldsymbol{\phi}$ を次のように定義する：

$$\mathbf{V}' = \left(\frac{n_1 - n\pi_1}{\sqrt{n\pi_1}}, \cdots, \frac{n_k - n\pi_k}{\sqrt{n\pi_k}}\right), \tag{6a.1.1}$$

$$\boldsymbol{\phi}' = (\sqrt{\pi_1}, \cdots, \sqrt{\pi_k}). \tag{6a.1.2}$$

われわれの目的は，$\mathbf{V}$ の（すなわち，確率変数とみなされる度数 $n_1, \cdots, n_k$ の）1次および2次の関数の漸近分布を調べることにある．これらについて次の結果が成り立つ．

（i）$\mathbf{b}$ を固定ベクトルとするとき，1次関数 $\mathbf{b}'\mathbf{V}$ の漸近分布は，平均零，分散 $\mathbf{b}'\mathbf{b} - (\mathbf{b}'\boldsymbol{\phi})^2 = \mathbf{b}'(\mathbf{I} - \boldsymbol{\phi}\boldsymbol{\phi}')\mathbf{b}$ の正規分布である．

いま離散的確率変数 X を考え，第 i 事象が生起したとき $b_i/\sqrt{\pi_i}$ という値をとるものとする．すなわち $P(X = b_i/\sqrt{\pi_i}) = \pi_i$ であるから

$$E(X)=\sum\frac{b_i}{\sqrt{\pi_i}}\pi_i=\sum b_i\sqrt{\pi_i}=\mathbf{b}'\boldsymbol{\phi}$$

$$V(X)=\sum\frac{b_i^2}{\pi_i}\pi_i-(\mathbf{b}'\boldsymbol{\phi})^2=\mathbf{b}'\mathbf{b}-(\mathbf{b}'\boldsymbol{\phi})^2$$

である．n 個の独立な観測があって，$X=b_i/\sqrt{\pi_i}$ $(i=1,\cdots,k)$ が n_i 回起こる．n 個の観測値の平均は

$$\bar{X}_n=\left(\frac{b_1}{\sqrt{\pi_1}}n_1+\cdots+\frac{b_k}{\sqrt{\pi_k}}n_k\right)\div n$$

である．中心極限定理［(i), **2c.5**］によって

$$\sqrt{n}(\bar{X}_n-\mathbf{b}'\boldsymbol{\phi})\xrightarrow{L}Y\sim N[0,\ \mathbf{b}'\mathbf{b}-(\mathbf{b}'\boldsymbol{\phi})^2]\tag{6a.1.3}$$

である．ここで $\sqrt{n}(\bar{X}_n-\mathbf{b}'\boldsymbol{\phi})=\mathbf{b}'\mathbf{V}$．よって上の結果を得る．

（ii）　p 個の1次関数 $\mathbf{B}'\mathbf{V}$，ここに $\mathbf{B}$ は $k\times p$ の行列で階数は p，の漸近分布は平均が零で分散共分散行列が $\mathbf{B}'(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{B}$ の多変量正規分布である．

この結果を確かめるには p 個の確率変数 $\mathbf{B}'\mathbf{V}$ の1次関数の漸近分布を調べるだけでよい．$\mathbf{V}$ について1次の $\boldsymbol{\lambda}'\mathbf{B}'\mathbf{V}$ を考えると，(i) の結果により

$$\boldsymbol{\lambda}'\mathbf{B}'\mathbf{V}\xrightarrow{L}Y\sim N[0,\ \boldsymbol{\lambda}'\mathbf{B}'(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{B}\boldsymbol{\lambda}]\tag{6a.1.4}$$

である．

$\mathbf{Z}$ を平均零，分散共分散行列 $\mathbf{B}'(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{B}$ の p 変量正規分布とする．1次関数 $\boldsymbol{\lambda}'\mathbf{Z}$ の分布は(6a.1.4) の右辺と同じである．したがって多変量中心極限定理［(iv), **2c.5**］より

$$\mathbf{B}'\mathbf{V}\xrightarrow{L}\mathbf{Z}\sim N_p[0,\ \mathbf{B}'(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{B}]\tag{6a.1.5}$$

である．ただし，N_p は p 変量正規分布をあらわす．（N_p の定義については **3b.3** を，またその詳細については第8章を参照.）

（iii）　$\mathbf{A}$ を $k\times(k-1)$ の行列とし，この $\mathbf{A}$ を含む $k\times k$ の行列 $(\boldsymbol{\phi}\mid\mathbf{A})$ は直交行列とする．このとき $k-1$ 個の1次関数 $\mathbf{G}=\mathbf{A}'\mathbf{V}$ の漸近分布は，それぞれ平均零，分散1をもつ $k-1$ 個の独立な正規変数の分布となる．

(ii) の結果と $\mathbf{A}'\boldsymbol{\phi}=0$ の条件を用いれば，$\mathbf{A}'\mathbf{V}$ の漸近分布の分散共分散行列は

$$\mathbf{A}'(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{A}=\mathbf{A}'\mathbf{A}=\mathbf{I}$$

で，$(k-1)\times(k-1)$ の行列となる．よって結果を得る．$\mathbf{G}\xrightarrow{L}Y\sim N_{k-1}(0,\mathbf{I})$.

（iv）　2次形式 $\mathbf{V}'\mathbf{C}\mathbf{V}$ の漸近分布が χ^2 分布になる十分条件は

$$\mathbf{C}^2=\mathbf{C}\quad\text{で}\quad\mathbf{C}\boldsymbol{\phi}=\alpha\boldsymbol{\phi}\tag{6a.1.6}$$

となることである．ここに α は定数であり，χ^2 の自由度は $\alpha=0$ なら $R(\mathbf{C})$ であり，$\alpha\neq 0$ ならば $[R(\mathbf{C})-1]$ である．ただし，$R(\mathbf{C})$ は $\mathbf{C}$ の階数を示す．

いま次の方程式を考える：

$$0=\boldsymbol{\phi}'\mathbf{V},\qquad\mathbf{G}=\mathbf{A}'\mathbf{V}$$

ここに $\mathbf{A}$ は (iii) と同じに定義される．$\mathbf{V}$ を解いて

$$\mathbf{V}=(\boldsymbol{\phi}\mid\mathbf{A})\begin{pmatrix}0\\ \hdashline\mathbf{G}\end{pmatrix}=\mathbf{A}\mathbf{G}$$

となる．これは $(\boldsymbol{\phi}\mid\mathbf{A})$ が直交行列であることによる．したがって

$$\mathbf{V}'\mathbf{C}\mathbf{V}=\mathbf{G}'\mathbf{A}'\mathbf{C}\mathbf{A}\mathbf{G}.$$

$\mathbf{V}'\mathbf{C}\mathbf{V}$ の漸近分布は $\mathbf{G}'\mathbf{A}'\mathbf{C}\mathbf{A}\mathbf{G}$ の漸近分布と同じで，$\mathbf{G}'\mathbf{A}'\mathbf{C}\mathbf{A}\mathbf{G}$ は $\mathbf{G}$ の連続関数である．したがってこの漸近分布は $\mathbf{G}$ の漸近分布から計算できる（(xi), **2c.4** 参照).

しかし $\mathbf{G}$ の漸近分布は $k-1$ 個の独立な正規分布で，それぞれ平均は零，分散は1である．$\mathbf{G}'\mathbf{A}'\mathbf{C}\mathbf{A}\mathbf{G}$ が χ^2 分布をする必要十分条件は $\mathbf{A}'\mathbf{C}\mathbf{A}$ がベキ等であること［(ii), **3b.4**]，すなわち

$$\mathbf{A}'\mathbf{C}\mathbf{A}\mathbf{A}'\mathbf{C}\mathbf{A}=\mathbf{A}'\mathbf{C}\mathbf{A}\tag{6a.1.7}$$

である．行列 $(\boldsymbol{\phi}\mid\mathbf{A})$ は直交しているから，

$$(\boldsymbol{\phi}\mid\mathbf{A})\begin{pmatrix}\boldsymbol{\phi}'\\ \hdashline\mathbf{A}'\end{pmatrix}=\boldsymbol{\phi\phi}'+\mathbf{A}\mathbf{A}'=\mathbf{I}$$

である．(6a.1.7) の $\mathbf{A}\mathbf{A}'$ をおきかえれば

$$\mathbf{A}'\mathbf{C}(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{C}\mathbf{A}=\mathbf{A}'\mathbf{C}\mathbf{A}\tag{6a.1.8}$$

となり，これは $(\boldsymbol{\phi}\mid\mathbf{A})$ が直交行列であるようなすべての $\mathbf{A}$ について成り立つ．(6a.1.8) が成り立つ十分条件は

$$\mathbf{C}^2=\mathbf{C}\quad\text{また}\quad\mathbf{C}\boldsymbol{\phi}=\alpha\boldsymbol{\phi}\tag{6a.1.9}$$

であること，すなわち $\mathbf{C}$ がベキ等であり，$\boldsymbol{\phi}$ が $\mathbf{C}$ の固有ベクトルであることである．χ^2 分布の自由度は $R(\mathbf{A}'\mathbf{C}\mathbf{A})$，すなわち $\mathbf{A}'\mathbf{C}\mathbf{A}$ の階数であることは **3b.4** の (ii) を用いればよい．したがって

$$\begin{aligned}R(\mathbf{C})&=R\left(\begin{pmatrix}\boldsymbol{\phi}'\\ \hdashline\mathbf{A}'\end{pmatrix}\mathbf{C}(\boldsymbol{\phi}\mid\mathbf{A})\right)=R\begin{pmatrix}\boldsymbol{\phi}'\mathbf{C}\boldsymbol{\phi}&\boldsymbol{\phi}'\mathbf{C}\mathbf{A}\\ \mathbf{A}'\mathbf{C}\boldsymbol{\phi}&\mathbf{A}'\mathbf{C}\mathbf{A}\end{pmatrix}\\&=R(\mathbf{A}'\mathbf{C}\mathbf{A}),\qquad\text{もし }\mathbf{C}\boldsymbol{\phi}=0\\&=R(\mathbf{A}'\mathbf{C}\mathbf{A})+1,\qquad\text{もし }\mathbf{C}\boldsymbol{\phi}=\alpha\boldsymbol{\phi},\ \alpha\neq 0\end{aligned}$$

となる．これが求める結果である．

（v）　$\mathbf{V}'\mathbf{C}\mathbf{V}$ の漸近分布が χ^2 であることの必要十分条件は

$$(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{C}(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{C}(\mathbf{I}-\boldsymbol{\phi\phi}')=(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{C}(\mathbf{I}-\boldsymbol{\phi\phi}')\tag{6a.1.10}$$

が成り立つことである．

(iv) では，必要十分条件は (6a.1.8) である．すなわち

$$\mathbf{A}'\mathbf{C}(\mathbf{I}-\boldsymbol{\phi\phi}')\mathbf{C}\mathbf{A}=\mathbf{A}'\mathbf{C}\mathbf{A}\tag{6a.1.11}$$

であることが示されている．ところで線形多様体 $\mathscr{M}(\mathbf{A})$ と $\mathscr{M}(\mathbf{I}-\boldsymbol{\phi\phi}')$ は同一であるか

ら，(6a. 1. 11) の $\mathbf{A}$ は $(\mathbf{I}-\boldsymbol{\phi}\boldsymbol{\phi}')$ でおきかえることができ，(6a. 1. 10) を得る．

(vi)　$\mathbf{V}'\mathbf{C}\mathbf{V}\xrightarrow{L} X\sim\chi^2(c)$，$\mathbf{V}'\mathbf{B}\mathbf{V}\xrightarrow{L} X\sim\chi^2(b)$ であり，$\mathbf{V}'\mathbf{D}\mathbf{V}$ が非負であるならば $\mathbf{B}=\mathbf{C}+\mathbf{D}\Rightarrow\mathbf{V}'\mathbf{D}\mathbf{V}\xrightarrow{L} X\sim\chi^2(b-c)$ となる．

3b. 4 の (iv) から得られる．

6a. 2　収　束　定　理

この章では変数ベクトル $(x_n, y_n, \cdots)$ あるいは $(\alpha_n x_n, \beta_n y_n, \cdots)$ の1次関数もしくは2次関数 h の漸近分布を考察する．ただし $\alpha_n\xrightarrow{P}\alpha$，$\beta_n\xrightarrow{P}\beta$ 等々とし，$\alpha,\beta,\cdots$ は定数であるとする．このような場合，2c. 4 の (x), (xiv) から，$(x_n, y_n, \cdots)$ が極限分布をもつならば

$$h(\alpha_n x_n, \beta_n y_n, \cdots)-h(\alpha x_n, \beta y_n, \cdots)\xrightarrow{P} 0$$

が成り立つことがわかる．したがって $h(\alpha_n x_n, \beta_n y_n, \cdots)$ の漸近分布と $h(\alpha x_n, \beta y_n, \cdots)$ の漸近分布は同じである．この結果の応用についていくつか考えてみよう．

(i)　いま (T_n) $(n=1,2,\cdots)$ を統計量の列とし

$$\sqrt{n}(T_n-\theta)\xrightarrow{L} X\sim N[0, \sigma^2(\theta)]$$

とする．さらに1変数の関数 g があり，1階の導関数 g' が存在するとしよう．このとき，$g'(\theta)\neq 0$ ならば

$$\sqrt{n}[g(T_n)-g(\theta)]\xrightarrow{L} X\sim N(0, [g'(\theta)\sigma(\theta)]^2) \tag{6a. 2. 1}$$

が成り立つ．さらに g' が連続とするならば

$$\frac{\sqrt{n}[g(T_n)-g(\theta)]}{g'(T_n)}\xrightarrow{L} X\sim N[0, \sigma^2(\theta)] \tag{6a. 2. 2}$$

が成り立つ．また $\sigma(\theta)$ も連続ならば

$$\frac{\sqrt{n}[g(T_n)-g(\theta)]}{g'(T_n)\sigma(T_n)}\xrightarrow{L} X\sim N(0,1) \tag{6a. 2. 3}$$

が成り立つ．

テイラー展開をして

$$g(T_n)-g(\theta)=(T_n-\theta)(g'(\theta)+\varepsilon_n)$$

を得るが，g についての仮定から $T_n\to\theta$ のとき，$\varepsilon_n\to 0$ である．これはどんな小さな量 e が与えられても $|T_n-\theta|<\delta$ である限り $|\varepsilon_n|<e$ とできることを意味する．したがって $n\to\infty$ のとき，仮定より $P(|\varepsilon_n|<e)\geqslant P(|T_n-\theta|<\delta)\to 1$ であり，e は任意であるから $\varepsilon_n\xrightarrow{P} 0$ である．$\sqrt{n}(T_n-\theta)$ は漸近分布をもち，$\varepsilon_n\xrightarrow{P} 0$ であるから，2c. 4 の (x)a を使って

$$\sqrt{n}[g(T_n)-g(\theta)]-\sqrt{n}(T_n-\theta)g'(\theta)=\sqrt{n}(T_n-\theta)\varepsilon_n\xrightarrow{P} 0$$

を得る．結局 $\sqrt{n}[g(T_n)-g(\theta)]$ の漸近分布は $\sqrt{n}(T_n-\theta)g'(\theta)$ の漸近分布と一致し，分布は $N(0, [g'(\theta)\sigma]^2)$ である．(6a. 2. 2) は $T_n\xrightarrow{P}\theta$ のとき $g'(T_n)\xrightarrow{P} g'(\theta)$ であるからただちに得られる．同様にして 2c. 4 の (x) を使って (6a. 2. 3) が出る．

標準誤差　$\sqrt{n}(T_n-\theta)$ の漸近分布が $N[0, \sigma^2(\theta)]$ のとき，通常 $\sigma(\theta)/\sqrt{n}$ を T_n の標準誤差という．標準誤差の推定値は $\sigma(T_n)/\sqrt{n}$ である．T_n は $N[\theta, \sigma^2(\theta)/n]$ と近似的に同じ分布に従うと表現してもよいであろう．また上の (6a. 2. 1) から (6a. 2. 3) までの結果から $g(T_n)$ の標準誤差は

$$g'(\theta)\times(T_n\text{ の標準誤差}) \tag{6a. 2. 4}$$

であることが示される．

注意　g の代りに n に陽に依存した関数 g_n を扱う場合には，(6a. 2. 1) から (6a. 2. 3) までの諸結果は一般には利用できない．そのような場合には，$g(T_n, n)$ の漸近分布を適当な標準化係数を使って求めることが必要になる．このために，g を2つの変数 (x, y) の関数と考え，$\partial g/\partial x$ は存在し，$y\to\infty, x\to\theta$ のとき極限 $G(\theta)\neq 0$ をもつと仮定する．g について仮定されたいくつかの条件の下で，一定の n について $T_n=\theta$ で $g(T_n, n)$ を展開すると，(i) の証明と同様にして

$$\frac{\sqrt{n}[g(T_n, n)-g(\theta, n)]}{h}\xrightarrow{L} X\sim N(0,1) \tag{6a. 2. 5}$$

を得る．ただし，h は次の形のうちのいずれでもよい．

(a)　$G(\theta)\sigma(\theta)$

(b)　$G(T_n)\sigma(T_n)$: G と σ が連続のとき

(c)　$(\partial g/\partial T_n)\sigma(T_n)$: σ が連続のとき

さらに，もし $n\to\infty$ で $\sqrt{n}[g(\theta, n)-f(\theta)]\to 0$ ならば $g(\theta, n)$ を $f(\theta)$ でおきかえることができる．

(ii)　$\mathbf{T}_n$ を k 次元統計量 $(T_{1n}, \cdots, T_{kn})$ とし，$\sqrt{n}(T_{1n}-\theta_1), \cdots, \sqrt{n}(T_{kn}-\theta_k)$ の漸近分布は平均0，分散共分散行列 $\Sigma=(\sigma_{ij})$ の k 変量正規分布であるとする．さらに，g を k 変数の関数とし全微分可能とする．このとき

$$\sqrt{n}u_n=\sqrt{n}[g(T_{1n}, \cdots, T_{kn})-g(\theta_1, \cdots, \theta_k)]$$

の漸近分布は平均0，分散 $v(\boldsymbol{\theta})$ $(\neq 0$ とする$)$

$$v(\boldsymbol{\theta})=\sum\sum\sigma_{ij}\frac{\partial g}{\partial\theta_i}\frac{\partial g}{\partial\theta_j} \tag{6a. 2. 6}$$

の正規分布である．

g は全微分可能であるから

$$g(T_{1n}, \cdots, T_{kn})-g(\theta_1, \cdots, \theta_k)=\sum(T_{in}-\theta_i)\left(\frac{\partial g}{\partial\theta_i}\right)+\varepsilon_n\|\mathbf{T}_n-\boldsymbol{\theta}\|$$

ただし $T_{in} \to \theta_i$ のとき $\varepsilon_n \to 0$ である．(i) の証明と同様にして $n \to \infty$ で $\varepsilon_n \xrightarrow{P} 0$ であり，$\sqrt{n}\|\mathbf{T}-\theta_n\| = [\sum n(T_{in}-\theta_i)^2]^{\frac{1}{2}}$ は漸近分布をもつから

$$\left|\sqrt{n}\,u_n - \sqrt{n}\sum(T_{in}-\theta_i)\frac{\partial g}{\partial \theta_i}\right| \xrightarrow{P} 0$$

である．ところで，$\sqrt{n}\sum(T_{in}-\theta_i)\partial g/\partial\theta_i$ は極限分布が正規分布の変数の1次関数であるから，これの漸近分布は平均0, 分散が (6a.2.6) で与えられる正規分布である．2c.4 の (ⅸ) により $\sqrt{n}\,u_n$ の漸近分布も同一である．

(6a.2.3) のときと同様に，σ_{ij} と g の偏導関数が θ の連続関数のときには

$$[\sqrt{n}\,u_n/\sqrt{v(\mathbf{T}_n)}] \xrightarrow{L} X \sim N(0,1) \qquad (6a.2.7)$$

ただし $v(\mathbf{T}_n)$ は $\theta = \mathbf{T}_n$ における $v(\theta)$ の値である．この結果を g がまた n の関数である場合に拡張する場合にはこの章の終りの問題9を参照せよ．

実際の応用の場合には，u_n 自身の分布が平均0, 分散の漸近値が v/n (v は (6a.2.6) で定義される) の漸近正規分布であるという言い方をする．また，混乱がないときには便宜のために添字の n を落としてもよい．

δ法　$g(T_{1n},\cdots,T_{kn})$ の標準誤差を求める方法は形式的に次のようにまとめることができる．g の全微分をとって

$$dg = \frac{\partial g}{\partial \theta_1}dT_{1n} + \cdots + \frac{\partial g}{\partial \theta_k}dT_{kn} \qquad (6a.2.8)$$

ここで dT_{in} は $T_{in}-\theta_i$ をあらわす．次に $T_{in}-\theta_i$ の分散，共分散が存在するものとして，分散を計算すると

$$V(dg) = \sum\sum\frac{\partial g}{\partial \theta_r}\frac{\partial g}{\partial \theta_s}\mathrm{cov}(T_{rn},T_{sn})$$
$$= \frac{1}{n}\sum\sum\frac{\partial g}{\partial \theta_r}\frac{\partial g}{\partial \theta_s}\sigma_{rs} \qquad (6a.2.9)$$

となる．

(ⅲ)　$\mathbf{T}_n$ を k 次元統計量 $(T_{1n},\cdots,T_{kn})$ とし，漸近分布は (ⅱ) と同じとする，$g_1,\cdots,g_q$ を k 変数の q 個の関数とし各 g_i は全微分可能とする．このとき

$$\sqrt{n}\,u_{in} = \sqrt{n}[g_i(T_{1n},\cdots,T_{kn}) - g_i(\theta_1,\cdots,\theta_k)] \qquad i=1,\cdots,q$$

の漸近分布は，平均0, 分散共分散行列が $\mathbf{G\Sigma G}'$ ($\mathbf{G} = (\partial g_i/\partial\theta_j)$) の q 変量正規分布である．分布の階数は $R(\mathbf{G\Sigma G}')$ に等しい (第8章参照)．

(ⅲ) の結果と実用上同等な表現は次のようになる．いま $(T_{1n},\cdots,T_{kn})$ が近似的に平均 $(\theta_1,\cdots,\theta_k)$, 分散共分散行列 $n^{-1}\mathbf{\Sigma}$ の k 変量正規分布に従うならば，$(g_1,\cdots,g_q)$ は近似的に平均 $[g_1(\theta),\cdots,g_q(\theta)]$, 分散共分散行列 $n^{-1}\mathbf{G\Sigma G}'$ の q 変量正規分布に従う．

(ⅲ) を証明するには，1次関数 $g = b_1g_1 + \cdots + b_qg_q$ を考え，(ⅱ) の結果を適用する

だけでよい．いま $\delta = b_1g_1(\theta) + \cdots + b_qg_q(\theta)$ とする．このとき $\sqrt{n}(g-\delta)$ の漸近分布は平均0, 分散

$$\sum\sum\frac{\partial g}{\partial \theta_r}\frac{\partial g}{\partial \theta_s}\sigma_{rs} = \sum\sum\left(\sum b_i\frac{\partial g_i}{\partial \theta_r}\right)\left(\sum b_i\frac{\partial g_i}{\partial \theta_s}\right)\sigma_{rs} = \mathbf{b'G\Sigma G'b} \qquad (6a.2.10)$$

をもつ正規分布である，ただし，$\mathbf{b}' = (b_1,\cdots,b_q)$ である．$\mathbf{b}$ は任意であるから求める結果が得られた．

(ⅲ) の結果は，分散共分散行列 $\mathbf{G\Sigma G}'$ がおそらく未知である母数に依存するから，実際の応用には非実用的である．そこで，母数 $\theta_1,\cdots,\theta_k$ の関数 σ_{ij} が連続である場合について，より強力な結果を以下の (ⅳ) で述べよう．

(ⅳ)　$\mathbf{G}_n, \mathbf{\Sigma}_n$ を $\theta_i = T_{in}$ $(i=1,\cdots,k)$ のときの $\mathbf{G}, \mathbf{\Sigma}$ の値とする．さらに，F_n を $\sqrt{n}\,u_{in}$ $(i=1,\cdots,q)$ の分布関数，H_n を平均0, 分散共分散行列 $\mathbf{G}_n\mathbf{\Sigma}_n\mathbf{G}_n'$ の q 変量正規分布の分布関数とする．このとき，$\mathbf{\Sigma}$ と $\mathbf{G}$ が θ について連続ならば

$$\lim_{n\to\infty}\sup_{x_1,\cdots,x_q}(F_n - H_n) = 0 \qquad (6a.2.11)$$

である．

(6a.2.11) の結果は，F_n と H_n の双方が同じ極限分布をもち，これらが変数に関して連続であることから導かれる．この結果から，F_n は q 変量正規分布で近似することができ，正規分布の分散共分散行列 $\mathbf{G}_n\mathbf{\Sigma}_n\mathbf{G}_n'$ は標本統計量のみを含んでいることがわかる．

推定値の併合　1つの母数の異なる場面から得られた独立な推定値を併合する問題に実際の場でわれわれはしばしば直面する．その場合には，各推定値の値とともに推定標準誤差が示されるのが普通である．このとき，2つの問題が考えられる．第1の問題は，すべての推定値が等質であるか，すなわち同じ値を推定しているのか，ということである．第2の問題は，もしすべての推定値が等質であるのならば，それらの推定値を組み合わせて1つの推定値を得るもっとも良いやり方はなにか，ということである．このような目標を念頭において，次のような命題を考えよう．

いま $T_1,\cdots,T_k$ を母数 $\theta_1,\cdots,\theta_k$ の独立な一致推定量とし，各標本の大きさを $n_1,\cdots,n_k$ とする．$\sqrt{n_i}(T_i-\theta_i)$ の漸近分布が，$n_i \to \infty$ で $N[0, s_i^2(\theta_i)]$ とする．次の統計量 H を (等質性の検定のために) 考えよう．

$$\mathrm{H} = \sum_{i=1}^{k}\frac{n_i(T_i-\hat\theta)^2}{s_i^2(T_i)} = \sum\frac{n_i}{s_i^2}T_i^2 - \left(\sum\frac{n_i}{s^2}T_i\right)^2\Big/\sum\frac{n_i}{s_i^2} \qquad (6a.2.12)$$

$$\hat\theta = \sum\frac{n_iT_i}{s_i^2(T_i)}\Big/\sum\frac{n_i}{s_i^2(T_i)}. \qquad (6a.2.13)$$

もし各 T_i が同じ量を推定しないのであれば，その差は統計量 (6a.2.12) に反映し，したがってそれを仮説 $\theta_1 = \cdots = \theta_k$ の検定に用いることができる．

(ⅴ)　$n = n_1 + \cdots + n_k$ とし，π_i は比 n_i/n $(i=1,\cdots,k)$ をあらわすものとする．

$$\mathrm{H} = \sum \frac{n_i(T_i - \hat{\theta})^2}{s_i^2(T_i)} = n \sum \frac{\pi_i(T_i - \hat{\theta})^2}{s_i^2(T_i)} \tag{6a. 2. 14}$$

の漸近分布は s_i $(i=1, \cdots, k)$ が連続関数であるという仮定と，仮説 $\theta_1 = \cdots = \theta_k$ の下で，$\pi_1, \cdots, \pi_k$ の値を固定して $n \to \infty$ とすると，自由度 $k-1$ のカイ2乗分布になる.

たとえば $\theta_1 = \cdots = \theta_k = \theta$ として

$$n \sum \frac{\pi_i(T_i - \hat{\theta})^2}{s_i^2(T_i)} = n \sum \frac{\pi_i(T_i - \theta)^2}{s_i^2(T_i)} - n(\theta - \theta)^2 \sum \frac{\pi_i}{s_i^2(T_i)} \tag{6a. 2. 15}$$

ただし,

$$\hat{\theta} = \sum \frac{\pi_i T_i}{s_i^2(T_i)} \Big/ \sum \frac{\pi_i}{s_i^2(T_i)} \tag{6a. 2. 16}$$

の分布を調べてみよう．s_i は連続であるから，次のことは容易にわかる.

$$n \sum \frac{\pi_i(T_i - \theta)^2}{s_i^2(T_i)} - n \sum \frac{\pi_i(T_i - \theta)^2}{s_i^2(\theta)} \xrightarrow{P} 0$$

$$n(\hat{\theta} - \theta)^2 \sum \frac{\pi_i}{s_i^2(T_i)} - n(\theta^* - \theta)^2 \sum \frac{\pi_i}{s_i^2(\theta)} \xrightarrow{P} 0$$

ここで θ^* は (6a. 2. 16) の $\hat{\theta}$ と同じで，$s_i^2(T_i)$ を $s_i^2(\theta)$ でおきかえたものである．H の漸近分布は,

$$\begin{aligned} \mathrm{H}' &= n \sum \frac{\pi_i(T_i - \theta)^2}{s_i^2(\theta)} - n(\theta^* - \theta)^2 \sum \frac{\pi_i}{s_i^2(\theta)} \\ &= n \sum \frac{(T_i - \theta)^2}{c_i^2} - n(\theta^* - \theta)^2 \sum \frac{1}{c_i} \end{aligned} \tag{6a. 2. 17}$$

の漸近分布と同じである．ただし $c_i^2 = s_i^2(\theta)/\pi_i$，$\theta^* = (\sum T_i/c_i^2)/\sum(1/c_i^2)$ である．いま

$$y_i = \sqrt{n}(T_i - \theta)/c_i$$

と書けば

$$\mathrm{H}' = \sum y_i^2 - \left(\sum \frac{y_i}{c_i}\right)^2 \Big/ \left(\sum \frac{1}{c_i^2}\right) \tag{6a. 2. 18}$$

となることがわかる．ただし y_i の漸近分布は $N(0,1)$ である．H′ は $y_1, \cdots, y_k$ に関して2次である（したがって連続でもある）から，H′ の漸近分布は 2次形式 (6a. 2. 18) で y_i が $N(0,1)$ に従いかつ互いに独立という仮説のもとでの分布と同じになる．3b. 4 の (ii) で述べた基準を利用すると，2次形式 (6a. 2. 18) が自由度 $k-1$ のカイ2乗分布をすることが容易にわかる.

H が大きくなく，あるいは一定の有意水準で有意でないならば,（自然な）併合推定値は

$$\hat{\theta} = \sum \frac{n_i T_i}{s_i^2(T_i)} \Big/ \sum \frac{n_i}{s_i^2(T_i)} \tag{6a. 2. 19}$$

で，推定漸近分散は

$$\left\{\sum \frac{n_i}{s_i^2(\hat{\theta})}\right\}^{-1} \tag{6a. 2. 20}$$

である．実用上，T_i を用いた式 (6a. 2. 14) は統計量 H を計算するのに有用である.

注意 (i) で示されたように $\sqrt{n_i}(T_i - \theta_i)$ の漸近分布が $N[0, s_i^2(\theta_i)]$ であるときには，$\sqrt{n_i}[g(T_i) - g(\theta_i)]$ の漸近分布は $N[0, u_i^2(\theta_i)]$ である．ただし u_i は s_i と g の導関数に依存する．もし g が1対1の関数であるならば，仮説 $\theta_1 = \cdots = \theta_k$ は $g(\theta_1) = \cdots = g(\theta_k)$ という仮説と同等である．そこで適当な関数 g を選び，$T_1, \cdots, T_k$ の代りに $g(T_1), \cdots, g(T_k)$ についての H 検定を用いる問題を扱うこととなる．同様に $g(\theta)$ の併合推定値を得ることができ，逆算により θ の併合推定値を得る．この問題には満足できる答はないが，$s_1^2 = \cdots = s_k^2$ のときには対応する u_i^2 も等しく，g も u_i が θ_i と独立なようにきめることができる（6g を参照）．そのように g を選ぶことは都合がよいであろう．H 検定の変換統計量への応用については 6g で考察しよう.

6b 多項分布のカイ2乗検定

6b. 1 単純仮説からのずれの検定

いま $n_1, \cdots, n_k$ を観測度数，$\pi_1, \cdots, \pi_k$ を多項分布の確率とする．仮説 $\pi_i = \pi_i^0$ $(i=1, \cdots, k$ で，π_i^0 は指定した値) の検定について，K. ピアソンは

$$\chi^2 = \sum \frac{(n_i - n\pi_i^0)^2}{n\pi_i^0} = \sum \frac{(\text{観測度数} - \text{期待度数})^2}{\text{期待度数}} \tag{6b. 1. 1}$$

という基準を示唆した (1900).

(i) (6b. 1. 1) で定義される χ^2 の漸近分布は自由度 $k-1$ のカイ2乗分布である.

これを証明するには (6a. 1. 6) の十分条件を用いる．(6b. 1. 1) の統計量は $\mathbf{V}'\mathbf{C}\mathbf{V}$ の2次形式で $\mathbf{C} = \mathbf{I}$ である．(6a. 1. 6) の条件

$$\mathbf{C}^2 = \mathbf{I}^2 = \mathbf{I} = \mathbf{C}, \qquad \mathbf{C}\boldsymbol{\phi} = \mathbf{I}\boldsymbol{\phi} = \boldsymbol{\phi}$$

は $\alpha = 1 \neq 0$ でみたされる．したがって，(6b. 1. 1) の漸近分布は自由度が $\mathrm{rank}\,\mathbf{I} - 1 = k-1$ のカイ2乗分布である．証明終り.

6b. 2 適合度についてのカイ2乗検定

適合度の良さを判断する問題を一般化すると，各級の確率がより少ない数の未知母数の特定の関数であらわされるか否かを検定する問題になる.

(i) 各級の確率が q 個の未知母数 $(\theta_1, \cdots, \theta_q) = \boldsymbol{\theta}'$ を含む特定の関数 $\pi_1(\boldsymbol{\theta}), \cdots, \pi_k(\boldsymbol{\theta})$ であらわされるとする．さらに (a) $\hat{\boldsymbol{\theta}}$ を (5c. 2. 6) の意味で $\boldsymbol{\theta}$ の有効推定量とし，(b) 各 $\pi_i(\boldsymbol{\theta})$ は θ_j，$j = 1, \cdots, q$ に関して1階(のみ)の連続な偏導関数をもつか，また

は，各 $\pi_i(\theta)$ が $\theta_1, \cdots, \theta_q$ の全微分可能な関数とし，(c) 大きさ $k \times q$ の行列 $M = (\pi_r^{-1/2}\partial\pi_r/\partial\theta_s)$ の階数は，θ が真の値のとき q であるとする．このとき

$$\chi^2 = \sum \frac{(n_i - n\hat{\pi}_i)^2}{n\hat{\pi}_i} = \sum \frac{(0 - E)^2}{E} \tag{6b.2.1}$$

の漸近分布は $\chi^2(k-1-q)$ である．ただし $\hat{\pi}_i = \pi_i(\hat{\theta})$.

(5c.2.6) の意味で有効な推定量が上の条件 (b) があれば存在することは **5e.2** の (iv) で示されていることを注意しておこう．したがって (6b.2.1) の漸近分布は1階偏導関数の存在および連続性の仮定のみを前提として，あるいは単に $\pi_i(\theta)$ の全微分可能性のみを前提として証明される．簡単のために証明では $X_n - Y_n \xrightarrow{P} 0$ を $X_n \overset{a}{=} Y_n$ (漸近的に等しい) と記す．いま

$$\mathbf{U}' = \left(\frac{n(\hat{\pi}_1 - \pi_1)}{\sqrt{n\pi_1}}, \cdots, \frac{n(\hat{\pi}_k - \pi_k)}{\sqrt{n\pi_k}}\right),$$

$$\mathbf{M} = \left(\frac{1}{\sqrt{\pi_r}}\frac{\partial\pi_r}{\partial\theta_s}\right), \text{ 大きさが } k \times q \text{ で階数が } q \text{ とする,}$$

$$\mathbf{D}' = [\sqrt{n}(\hat{\theta}_1 - \theta_1), \cdots, \sqrt{n}(\hat{\theta}_q - \theta_q)],$$

$$\mathbf{Z} = \mathbf{M}'\mathbf{V}, \ \mathbf{V} \text{ は (6a.1.1) で定義される,}$$

としよう．また，情報行列を $(i_{rs}) = \mathbf{M}'\mathbf{M} = \boldsymbol{\varphi}$ としよう．$\mathbf{M}$ の階数が q であるという仮定から $\boldsymbol{\varphi}$ は θ が真の値のとき正則であり，また $\hat{\theta}$ は (5c.2.6) の意味で有効推定量であるから

$$\mathbf{D} \overset{a}{=} \boldsymbol{\varphi}^{-1}\mathbf{M}'\mathbf{V} \tag{6b.2.2}$$

である．条件 (b) を用いて $\pi_i(\hat{\theta})$ を θ のまわりで展開し

$$\frac{\sqrt{n}[\pi_i(\hat{\theta}) - \pi_i(\theta)]}{\sqrt{\pi_i(\theta)}} \overset{a}{=} \frac{1}{\sqrt{\pi_i(\theta)}}\sum_{s=1}^{q}\frac{\partial\pi_i}{\partial\theta_s}\sqrt{n}(\hat{\theta}_s - \theta_s) \tag{6b.2.3}$$

を得る．(6b.2.3) で $i = 1, \cdots, k$ とするときの関係式は行列記号を用いて $\mathbf{U} \overset{a}{=} \mathbf{MD}$ と書ける．ここで $\mathbf{D}$ を (6b.2.2) でおきかえると $\mathbf{U} \overset{a}{=} \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}'\mathbf{V}$ となる．次に $\mathbf{V}$ からこの両辺をそれぞれ引いて

$$\mathbf{V} - \mathbf{U} \overset{a}{=} (\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}')\mathbf{V} \tag{6b.2.4}$$

が求められる．(6b.2.1) の χ^2 統計量は分母の $\hat{\pi}_i$ を π_i でおきかえると，

$$\begin{aligned}(\mathbf{V} - \mathbf{U})'(\mathbf{V} - \mathbf{U}) &\overset{a}{=} \mathbf{V}'(\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}')(\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}')\mathbf{V} \\ &= \mathbf{V}'(\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}')\mathbf{V} \qquad (\mathbf{M}'\mathbf{M} = \boldsymbol{\varphi} \text{ だから})\end{aligned}$$

となる．カイ2乗分布をするための十分条件 (6a.1.6) は

$$(\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}')^2 = \mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}' \qquad (\mathbf{M}'\mathbf{M} = \boldsymbol{\varphi} \text{ だから})$$

$$(\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}')\boldsymbol{\phi} = \boldsymbol{\phi} \qquad (\mathbf{M}'\boldsymbol{\phi} = 0 \text{ だから})$$

であるから $\alpha = 1 \neq 0$ でみたされる．さらに，ベキ等性を用いて

$$\begin{aligned}\mathrm{rank}(\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}') &= \mathrm{trace}(\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}') \\ &= k - \mathrm{trace}\,\mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}' \\ &= k - \mathrm{trace}\,\boldsymbol{\varphi}^{-1}\mathbf{M}'\mathbf{M} \\ &= k - \mathrm{trace}\,\mathbf{I}_q = k - q\end{aligned}$$

である．ただし $\mathbf{I}_q$ は次数 q の単位行列．ゆえに **6a.1** の (iv) の結果を用いて (6b.2.1) の漸近分布は $\chi^2(k-q-1)$ であり，したがって $(\partial\pi_i/\sqrt{\pi_i}\partial\theta_j)$ の階数が q ならば自由度として $k-1-$(推定された母数の個数) と考えればよい．

注意　(6b.2.1) の漸近分布を求めるのに用いた条件の1つは，その期待度数を得るのに使われた $\hat{\theta}$ が有効推定量であることである．第5章で有効推定量の存在条件およびこれらの推定量の計算のしかたについて論じた．有効推定量としては広い範囲のものがあるが，そのいずれについても (6b.2.1) の漸近分布は適用可能である．しかし，最尤推定量が有効統計量であるという条件がみたされるときには，最尤推定量を用いるほうに有利な点がある (Rao (1961a, 1962d) 参照).

6b.3　1つの級における偏差の検定

カイ2乗による適合度の検定は，観測度数とモデルからきまる期待度数との間の差の検定に関する総合的な検定である．ある級の観測度数の期待度数からのずれが観測度数の不足によるものかどうかを調べることに興味があることもあろう．たとえば，事故の度数分布を調べるときには，無事故の人の数を少なめに見ることが多そうである．Cochran (1954) は j 番目の級における偏差を吟味するのに残差

$$r_j = \frac{n_j - n\pi_j(\hat{\theta})}{\sqrt{n\pi_j(\hat{\theta})}} \tag{6b.3.1}$$

を検定の基準とすることを提案した．統計量 (6b.3.1) の漸近分布を求めるには，漸近分散，すなわち漸近分布の分散を計算すれば十分である．

(6b.2.4) から

$$\mathbf{V} - \mathbf{U} \overset{a}{=} (\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}')\mathbf{V}, \qquad \mathbf{U} \overset{a}{=} \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}'\mathbf{V}$$

となることがわかる．$\mathbf{V} - \mathbf{U}$ と $\mathbf{U}$ とにそれぞれ漸近的に等しい上式の右辺の間の共分散は

$$\begin{aligned}&(\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}')(E(\mathbf{V}\mathbf{V}'))\mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}' \\ &\quad = (\mathbf{I} - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}')(\mathbf{I} - \boldsymbol{\phi}\boldsymbol{\phi}')\mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}' = 0 \qquad (\boldsymbol{\phi}'\mathbf{M} = 0 \text{ だから})\end{aligned}$$

となる．よって $\mathbf{V} - \mathbf{U}$ の各要素は $\mathbf{U}$ の全要素と漸近的に無相関である．これは重要な結果であり，最小2乗法で最良推定量と残差とは無相関であるという結果 [(ii), **4a.5**] と同様である．

よって $\mathbf{V}-\mathbf{U}$ と $\mathbf{U}$ の j 番目の要素の間の漸近共分散は零となり，

$$\text{漸近共分散}\left\{\frac{n_j-n\pi_j(\hat{\theta})}{\sqrt{n\pi_j(\hat{\theta})}},\ \frac{n\pi_j(\hat{\theta})-n\pi_j(\theta)}{\sqrt{n\pi_j(\hat{\theta})}}\right\}=0$$

すなわち

$$\text{漸近分散}\left\{\frac{n_j-n\pi_j(\theta)}{\sqrt{n\pi_j(\hat{\theta})}}\right\}=\text{漸近分散}\left\{\frac{n_j-n\pi_j(\hat{\theta})}{\sqrt{n\pi_j(\hat{\theta})}}\right\}+\text{漸近分散}\left\{\frac{n\pi_j(\hat{\theta})-n\pi_j(\theta)}{\sqrt{n\pi_j(\hat{\theta})}}\right\}$$

を得る．ところで

$$\text{漸近分散}\left\{\frac{n_j-n\pi_j(\theta)}{\sqrt{n\pi_j(\hat{\theta})}}\right\}=\text{分散}\left\{\frac{n_j-n\pi_j(\theta)}{\sqrt{n\pi_j(\theta)}}\right\}=1-\pi_j(\theta)$$

であり，(6a.2.6) を使えば

$$\text{漸近分散}\left\{\frac{n\pi_j(\hat{\theta})-n\pi_j(\theta)}{\sqrt{n\pi_j(\hat{\theta})}}\right\}=\text{漸近分散}\left\{\frac{n\pi_j(\hat{\theta})-n\pi_j(\theta)}{\sqrt{n\pi_j(\theta)}}\right\}=\sum_s\sum_t\frac{1}{\pi_j}\frac{\partial\pi_j}{\partial\theta_t}\frac{\partial\pi_j}{\partial\theta_s}i^{st}$$

である．ゆえに

$$\text{漸近分散}\left\{\frac{n_j-n\pi_j(\hat{\theta})}{\sqrt{n\pi_j(\hat{\theta})}}\right\}=v(\theta)=1-\pi_j(\theta)-\sum_s\sum_t\frac{1}{\pi_j}\frac{\partial\pi_j}{\partial\theta_s}\frac{\partial\pi_j}{\partial\theta_t}i^{st}\qquad(6b.3.2)$$

を得る．第 j 番目の級の偏差についての正規偏差量 $N(0,1)$ に基づく検定は

$$\frac{n_j-n\pi_j(\hat{\theta})}{[nv(\hat{\theta})\pi_j(\hat{\theta})]^{\frac{1}{2}}}\qquad(6b.3.3)$$

となる．

母数が1つのとき (6b.3.2) は

$$v(\theta)=(1-\pi_j)-\frac{1}{\pi_j}\left(\frac{d\pi_j}{d\theta}\right)^2\frac{1}{i(\theta)}\qquad(6b.3.4)$$

となる．ただし $i(\theta)$ は観測値1個当りの情報量である．

例として，ポアソン分布に従う変量の零の級における偏差について考えてみよう．そこでは

$$\pi_0=e^{-\mu},\qquad\frac{d\pi_0}{d\mu}=-e^{-\mu},\qquad i=\frac{1}{\mu}$$

である．(6b.3.4) の値は

$$1-e^{-\mu}-\mu e^{-\mu}$$

となる．観測度数分布から次の数字を得たとしよう．

$$n=200,\qquad n_0=15,\qquad \hat{\mu}\ (\mu\text{ の推定値})=2.1055$$

$$\begin{aligned}n_0-n\hat{\pi}_0&=15-200e^{-\hat{\mu}}\\&=15-24.356=-9.356\\n\hat{\pi}_0v(\hat{\mu})&=ne^{-\hat{\mu}}(1-e^{-\hat{\mu}}-\hat{\mu}e^{-\hat{\mu}})\\&=24.356\times(0.62181)=15.1448.\end{aligned}$$

(6b.3.3) の値は絶対値で

$$\frac{9.356}{\sqrt{15.1448}}=2.40$$

となり，正規変数としては大きな値で零の級における観測度数の不足を示している．

6b.4　母数がある部分集合に含まれるか否かの検定

$\pi_i(\theta)$ の指定が正しいものとして，θ が，軌跡

$$\theta_i=g_i(\tau_1,\cdots,\tau_r),\qquad i=1,\cdots,q\ ;\qquad r<q\qquad(6b.4.1)$$

によって規定される θ の許容値の部分集合に属するかどうかを調べてみよう．ただし，$\tau_1,\cdots,\tau_r$ は新しい母数，g_i はそれらについて連続な1階偏導関数をもつとする．(6b.4.1) の仮説を検定するために，著者は次の検定基準を提案した (Rao (1961a))．すなわち

$$\chi^2=\sum\frac{(E_1-E_2)^2}{E_2}\qquad(6b.4.2)$$

ここで E_1 と E_2 とはそれぞれ θ に制限がない場合の推定度数と (6b.4.1) の制限がある場合の推定度数とを示す．

さらに特殊の場合には，$\hat{\theta}$ は制限がない場合の θ の有効推定値とする．仮説 (6b.4.1) が正しいときには π_i は $\tau'=(\tau_1,\cdots,\tau_r)$ の関数となる．$\hat{\tau}$ を τ の有効推定量とすると

$$\chi^2=\sum\frac{[n\pi_i(\hat{\theta})-n\pi_i(\hat{\tau})]^2}{n\pi_i(\hat{\tau})}\qquad(6b.4.3)$$

の漸近分布は，$\pi_i(\theta)$ が θ の関数としてみたす条件と同様の条件を $\pi_i(\tau)$ がみたすならば，$\chi^2(q-r)$ の分布と一致する．

この証明は 6b.2 と同じようなやり方でできる．ここで，τ と θ にたいする情報行列 φ_1 と φ とが $\varphi_1=\Delta'\varphi\Delta$ という関係で結ばれていることを注意しておく．ただし $\Delta=(\partial\theta_i/\partial\tau_j)$ は導関数の行列とする．$\sqrt{n}[\pi_i(\hat{\theta})-\pi_i(\hat{\tau})]$ は $\sqrt{n}[\pi_i(\hat{\theta})-\pi_i(\theta)]$ の場合と同様に観測度数の項で展開できる．したがって，統計量 (6b.4.3) は $\mathbf{V}$ の2次形式と漸近的に等しくなる．そこで (6a.1.6) を使えば結果が得られる．

6b.5　例　　題

例 1　ベーツソンは，スイートピーの紫と赤の花の色と花粉の形の長いのと丸いものを制御する2つの遺伝子の分離に関して次のデータを示している．

その結果は，独立に分離するという仮説の下では，その期待度数の比率が 9：3：3：1 になるような交雑から得られたものである．データはその期待度数と一致しているか？ 6b.1 の χ^2 検定を適用することができる．

	紫-長い	赤-長い	紫-丸い	赤-丸い	計
観測度数	296	27	19	85	427
期待度数	$3843 \div 16$	$1281 \div 16$	$1281 \div 16$	$427 \div 16$	427
$\dfrac{(観測度数)^2}{期待度数}$	364.7817	9.1054	4.5090	270.7260	649.1221

$$\chi^2 = \sum \frac{(観測度数)^2}{期待度数} - \sum(観測度数) = 649.1221 - 427 = 222.1221 \quad (自由度3)$$

自由度3の χ^2 が222.1221を越える確率は非常に小さいから，期待度数からのずれがあることを示している．大標本の検定においては，各区画の期待度数は5よりも大きいことが必要であるといえる．期待度数の小さい場合の正確な検定は例2で示すようなやり方による．これ以外の方法については**6d.2**および**6d.3**で考察する．

例 2　観測度数が8, 3, 1となったとき，等確率の3項分布から起こったものかどうかを検定しよう．

期待度数4, 4, 4はすべて小さいから，χ^2 近似は正しくないかもしれない．もしこのことを無視すると，自由度2の χ^2 は

$$\frac{8^2}{4} + \frac{3^2}{4} + \frac{1^2}{4} - 12 = 6.5$$

となり，2%と5%の間の確率となる．この例では期待度数はすべて小さいから χ^2 近似が正しいかどうかは疑わしい．こう疑ってみて，χ^2 の値が6.5に等しいかあるいは越える実際の確率を評価してみる．観測度数 x, y, z が得られる確率は

$$\frac{12!}{x!y!z!}\left(\frac{1}{3}\right)^{12}$$

である．

総度数12の分割のしかたは全部で91通りあり，このうち $\chi^2 \geqslant 6.5$ となる分割の確率を計算すると次のようになる：

分割の型	個数	確率	χ^2
12　0　0	3	0.0^51882	24
11　1　0	6	0.0^42258	18.5
10　2　0	6	0.0^31242	14
10　1　1	3	0.0^32484	13.5
9　3　0	6	0.0^34140	10.5
9　2　1	6	0.0^21242	9.5
8　4　0	6	0.0^39314	8
8　3　1	6	0.0^23726	6.5
7　5　0	6	0.0^21490	6.5

χ^2 が6.5以上となる正確な確率は，これらの確率を加えて0.04844と求まるが，これは漸近分布から求めた値よりもいく分大きな値である．一般にはこのように近い値となるとは限らない．ただし，各級内の期待度数が5よりも大きい場合にはきわめて近い値となる．

6b.6　いくつかの級における偏差の検定

6b.2で定義した **U**, **M**, **6a.1**で定義した **V** のほかに

$$\mathbf{W}' = \left(\frac{n_1 - n\pi_1(\hat{\theta})}{\sqrt{n\pi_1(\hat{\theta})}}, \quad \cdots, \quad \frac{n_k - n\pi_k(\hat{\theta})}{\sqrt{n\pi_k(\hat{\theta})}}\right) \tag{6b.6.1}$$

を定義する．これは k 個の級における偏差をあらわすベクトルである．これを用いれば**6b.2**で述べた適合度の χ^2 検定は $\mathbf{W}'\mathbf{W}$ ともあらわされる．

6b.3では1つの級における偏差を調べる検定を述べた．ここでは，r 個の指定した級における偏差の検定について述べる．このために，$\mathbf{U} - \mathbf{V}$ と漸近的に等しい **W** の漸近分散共分散行列 (a.D) を求めると，

$$\mathrm{a.D}(\mathbf{W}) = \mathrm{a.D}(\mathbf{V} - \mathbf{U}) = \mathrm{a.D}(\mathbf{V}) - \mathrm{a.D}(\mathbf{U}) \tag{6b.6.2}$$

となる．これは $\mathbf{V} - \mathbf{U}$ と **U** との漸近共分散が**6b.3**で示したように零になるからである．さらに

$$\mathrm{a.D}(\mathbf{V}) - \mathrm{a.D}(\mathbf{U}) = (\mathbf{I} - \boldsymbol{\phi}\boldsymbol{\phi}') - \mathbf{M}\boldsymbol{\varphi}^{-1}\mathbf{M}' \tag{6b.6.3}$$

である．

その偏差を検定すべき級の部分集合に対応する項のみからなる **W** の部分ベクトルを **d** で，(6b.6.3)で **d** に関係する要素だけを残して得られる **d** の分散共分散行列を **B** とする．**B** の要素は未知の θ を含んでいる．θ を $\hat{\theta}$ でおきかえ，行列 $\hat{\mathbf{B}}$ が得られる．r 個の特定の級における偏差を吟味するためには，**B** の階数が r のとき

$$\chi^2 = \mathbf{d}'\hat{\mathbf{B}}^{-1}\mathbf{d} \tag{6b.6.4}$$

を自由度 r の χ^2 で検定する．$R(\mathbf{B}) = r$ でないときには，$\hat{\mathbf{B}}$ の一般逆行列を用い，χ^2 の自由度は **B** の階数と考えねばならない．

6c　多項分布からの独立標本に関する検定

6c.1　一般的な結果

m 個の多項母集団から，それぞれ大きさ $n_{1\cdot}, \cdots, n_{m\cdot}$ の標本を得たものとする．ここで m 個の多項母集団の級の数は必ずしも等しくなくてよいとする．1つの多項母集団の場合と同様に，2種類の問題が生じる．1つは，モデルの規定（すなわち，各級の確率をいく

つかの母数の関数として規定できるという）をすべての多項分布について同時に検定することである．第2は，それらの母数が許容値の集合の部分集合に属するかどうかを検定することである．

たとえば，2つの異なる母集団から血液型 O, A, B, AB の度数を調べたとする．第1には，各血液型の度数が第5章の **5g** で論じたベルンシュタインの仮説と一致するかどうかを調べることができる．当然のこととして起こる第2の問題は，遺伝子の度数が2つの母集団で同じかどうかということである．p と q をそれぞれ第1の母集団での A, B の遺伝子の度数とし，p' と q' とを第2の母集団での遺伝子の度数とする．仮説は $p=p'$, $q=q'$ ということを述べている．

次に，第 i 母集団の各級の確率が

$$\pi_{i1}(\theta),\quad \cdots,\quad \pi_{ik_i}(\theta) \tag{6c.1.1}$$

と q 個の未知母数 $\theta_{1i},\cdots,\theta_{qi}$ の関数として規定される場合の一般的問題を考察しよう．$n_{i1},\cdots,n_{ik_i}$（$\sum n_{ij}=n_{i.}$）を第 i 標本の観測度数とし，$\hat{\theta}_{1i},\cdots,\hat{\theta}_{qi}$ を第 i 標本に基づく最尤推定量とする．いま (6c.1.1) が **6b.2** の χ^2 適合度検定の適用条件をみたすものとする．このとき，第 i 標本の適合度検定の χ^2 は

$$\chi_i^2=\sum_{j=1}^{k_i}\frac{(n_{ij}-n_{i.}\hat{\pi}_{ij})^2}{n_{i.}\hat{\pi}_{ij}} \tag{6c.1.2}$$

で自由度は k_i-q-1 である．ここで $\hat{\pi}_{ij}$ は π_{ij} の未知母数を $\hat{\theta}_{1i},\cdots,\hat{\theta}_{qi}$ でおきかえたものである．m 個の標本からの χ^2（すなわち，個別の適合度 χ^2 の m 個の合計）は

$$\sum_{i=1}^{m}\chi_i^2=\sum_{i=1}^{m}\sum_{j=1}^{k_i}\frac{(n_{ij}-n_{i.}\hat{\pi}_{ij})^2}{n_{i.}\hat{\pi}_{ij}} \tag{6c.1.3}$$

であり，自由度は $\sum k_i-mq-m$ で，これが全確率の規定についての総合的な検定のための統計量となる．

いますべての r および i について $\theta_{ri}=\theta_r$，すなわち m 個の母集団について母数は共通であるとする．また，$\theta_1^*,\cdots,\theta_q^*$ を m 個の標本から求めた $\theta_1,\cdots,\theta_q$ の最尤推定量，π_{ij}^* を $\pi_{ij}(\theta)$ の $\theta_1,\cdots,\theta_q$ を $\theta_1^*,\cdots,\theta_q^*$ でおきかえたものとする．このとき，観測度数 n_{ij} の期待値の推定量として次の2つ

$$E_1=(\text{期待値})_1=n_{i.}\hat{\pi}_{ij},$$
$$E_2=(\text{期待値})_2=n_{i.}\pi_{ij}^*$$

が得られる．ここに $\hat{\pi}_{ij}$ は (6c.1.2) と同じように第 i 標本のみに基づく最尤推定量でおきかえたものであることを注意しておく．母数がすべての標本について共通であるという仮説の χ^2 検定は

$$\sum\sum\frac{(E_1-E_2)^2}{E_2}=\sum_{i=1}^{m}\sum_{j=1}^{k_i}\frac{(n_{i.}\hat{\pi}_{ij}-n_{i.}\pi_{ij}^*)^2}{n_{i.}\pi_{ij}^*} \tag{6c.1.4}$$

が漸近的に $\chi^2[q(m-1)]$ となることを用いる．証明は (6b.4.2) と同様である．

6c.2　類似標本の一様性の検定

2つの標本の k 個の階級の度数が表 6c.2α のように得られたとする．

表 6c.2α　2つの類似標本における観測度数

標本	階級 (1)	(2)	…	(k)	計
第1組	n_{11}	n_{12}	…	n_{1k}	$n_{1.}$
第2組	n_{21}	n_{22}	…	n_{2k}	$n_{2.}$
計	$n_{.1}$	$n_{.2}$	…	$n_{.k}$	$n_{..}$
第1組/計	p_1	p_2	…	p_k	p

ここで階級というのは，離散的な分類でも，連続変数の区間であってもよい．もし各区画の確率について仮説（すなわち，より少ない数のパラメータの関数として）がなにも規定していないときは，2つの標本が同一の母集団からとられたものであるということはどのようにして検定することができるか？

いま $\pi_{11},\cdots,\pi_{1k}$ および $\pi_{21},\cdots,\pi_{2k}$ を2つの母集団の各区画の確率とする．検定すべき仮説は $\pi_{1i}=\pi_{2i}$, $i=1,\cdots,k$ である．第1の標本から π_{1i} の推定値を求めると

$$\hat{\pi}_{1i}=\frac{n_{1i}}{n_{1.}}$$

であり，第2の標本からの π_{2i} の推定値は

$$\hat{\pi}_{2i}=\frac{n_{2i}}{n_{2.}}$$

となる．いま，$\pi_{1i}=\pi_{2i}=\pi_i$（とおく）とすると，この共通の値の推定値は（2つの標本を一緒にして）

$$\hat{\pi}_i=\frac{n_{1i}+n_{2i}}{n_{..}}=\frac{n_{.i}}{n_{..}}$$

となる．検定すべき仮説に対する (6c.1.4) の χ^2 は

$$\sum\sum\frac{(E_1-E_2)^2}{E_2}=\sum_{i=1}^{k}\frac{\left(n_{1i}-\frac{n_{1.}n_{.i}}{n_{..}}\right)^2}{\frac{n_{1.}n_{.i}}{n_{..}}}+\sum_{i=1}^{k}\frac{\left(n_{2i}-\frac{n_{2.}n_{.i}}{n_{..}}\right)^2}{\frac{n_{2.}n_{.i}}{n_{..}}}$$

$$=\frac{1}{p(1-p)}\sum n_{.i}(p_i-p)^2=\frac{1}{p(1-p)}\left(\sum n_{.i}p_i^2-n_{..}p^2\right)$$

$$=\frac{1}{p(1-p)}\left(\sum n_{1i}p_i-n_{1.}p\right) \tag{6c.2.1}$$

となる．ここで p_i と p は表 6c.2α で定義されているものである．仮説によって $k-1$ 個の独立な区画の確率が2標本で等しいことが規定されているから，(6c.2.1) で定義される χ^2 の自由度は $k-1$ である．

一般に，m 個の標本が比較されるときには，その式は

$$\chi^2 = \sum\sum \frac{\left(n_{ij} - \frac{n_i.n._j}{n..}\right)^2}{\frac{n_i.n._j}{n..}}, \qquad 自由度 = (m-1)(k-1) \qquad (6c.2.2)$$

となる．一般の場合には，χ^2 の式はこれ以上簡単にはできない．

6c.3 例　題

集団 C に属する 353 人と集団 D に属する 364 人の，4 つの血液型 O, A, B, AB についての人数の分布が表 6c.3α に与えられている．各血液型の比率が 2 つの集団で等しいという仮説を吟味するために (6c.2.1) の検定を適用する．表 6c.3α で計算した $\chi^2 = 11.77$, 自由度 3 は 5％水準で有意であるから，2 つの集団には差があることがわかる．

表 6c.3α　血液型による個人の分類

	O	A	B	AB	計
集　団C	121	120	79	33	353
集　団D	118	95	121	30	364
計	239	215	200	63	717
$p = \frac{集団C}{計}$	0.5063	0.5581	0.3950	0.5238	0.4923

$$\sum n_{1i}p_i - n_1.p = 121(0.5063) + 120(0.5581) + \cdots - 353(0.4923) = 2.9428$$
$$p(1-p) = 0.24994, \quad \chi^2 = 2.9428/0.24994 = 11.77, \quad 自由度 3$$

しかし，目的が集団の差を証明することにあるのなら，自由度 3 の χ^2 検定は利用できるデータにとって最良のものではない．2 つの集団間に内在的な（つまり遺伝的な）差がもしあるならば，それは遺伝子 O, A, B の相対度数に現れる．遺伝子の度数が等しいという仮説の検定は，表現型の度数が遺伝子の度数の既知の関数であることを認識しないで，表現型の度数が等しいという仮説を検定することよりもより基本的である．遺伝子の度数のようなより内在的な特性へのそのような関数的な従属関係が未知，または疑わしいときにかぎり，表現型の度数を含む (6c.2.1) 型の検定が正当化される．

表 6c.3β　最尤推定値

	$\hat{p}$	$\hat{q}$	$\hat{r} = (1-\hat{p}-\hat{q})$
集　団C	0.24649	0.17317	0.58034
集　団D	0.19025	0.23573	0.57401
合併集団	0.21762	0.20435	0.57803

表現型 O, A, B, AB の確率を遺伝子の度数 p, q, r であらわすやり方は，5g の例題で示した．その節で示した最尤法を用いると，第 1 の標本，第 2 の標本および合併標本の p, q, r の推定値が得られる（表 6c.3β）．これらの推定値を用いて，各推定値およびそれらの結合推定値に基づく各集団の期待度数（E_1 および E_2）が求められ，表 6c.3γ に示される．

表 6c.3γ　2 つの仮説のもとでの期待度数

型	確　率	集　団C			集　団D		
		観測度数	E_1	E_2	観測度数	E_1	E_2
O	r^2	121	118.89	117.94	118	119.93	121.62
A	$p^2 + 2pr$	120	122.44	105.52	95	92.68	108.81
B	$q^2 + 2qr$	79	81.54	98.14	121	118.74	101.19
AB	$2pq$	33	30.13	31.40	30	32.65	32.37
計	1	353	353	353	364	364	364

χ^2（適合度検定）

$$\chi^2(集団C) = \sum \frac{(O-E_1)^2}{E_1} = 0.44 \qquad 自由度 1$$
$$\chi^2(集団D) = \sum \frac{(O-E_1)^2}{E_1} = 0.35 \qquad 自由度 1$$
$$\chi^2\ (集団の一様性) = \sum\sum \frac{(E_1-E_2)^2}{E_2} = 11.04 \qquad 自由度 2$$

個別集団に関する χ^2 の値は小さいから，区画の確率を遺伝子度数によって規定するのは正しい．最後の χ^2 の値は，表現型度数の直接比較によって得られた自由度 3 の χ^2 の値 11.77 よりも，2 つの集団の差をよりよく証明している．2 つの χ^2 の値はほとんど同じ大きさであるが，遺伝子度数の比較に基づく χ^2 のほうが自由度は 1 小さいから，帰無仮説を支持する確率はそれだけ小さい．

この検定はいくつかの集団の差を同時に調べる場合に拡張できる．その場合には標本の各組および合併した標本について最尤推定値を求める必要がある．そのうえで，2 組の期待度数を計算して χ^2 検定によって比較する．

6d 分　割　表

6d.1 観測結果の得られる確率と大標本の検定

母集団の各個体は，属性 A については r 個のカテゴリー $A_1, \cdots, A_r$ のどれか 1 つに属し，属性 B については s 個のカテゴリー $B_1, \cdots, B_s$ の 1 つに属する，等々というように

記載されているときには，その代表的な階級が $A_iB_j\cdots$ とあらわされるような $r\times s\times\cdots$ 個の階級における個体の度数分布が得られる．k 個の属性があるとき，いま述べた度数の配列は k 重の分割表とよばれる．この節で考える分割表は，$r\times s\times\cdots$ 個の階級をもつ多項分布の特別の場合にすぎないから，その分布については新しい問題はなにも起こらず，**6b** の一般的な結果が適用される．

本節では属性が 2 個ある場合のいろいろな問題について考えるが，一般の場合も扱いは同様である．階級 A_iB_j の観測度数を n_{ij} で示し，その確率を π_{ij} であらわす．また

$$n_{i1}+n_{i2}+\cdots+n_{is}=n_{i\cdot}\qquad \pi_{i1}+\pi_{i2}+\cdots+\pi_{is}=\pi_{i\cdot}$$

$$n_{1j}+n_{2j}+\cdots+n_{rj}=n_{\cdot j}\qquad \pi_{1j}+\pi_{2j}+\cdots+\pi_{rj}=\pi_{\cdot j}$$

$$n_{1\cdot}+n_{2\cdot}+\cdots=n_{\cdot 1}+n_{\cdot 2}+\cdots=n_{\cdot\cdot}$$

$$\pi_{1\cdot}+\pi_{2\cdot}+\cdots=\pi_{\cdot 1}+\pi_{\cdot 2}+\cdots=1$$

とする．この観測度数の得られる確率は，階級が $r\times s$ 個あるときの多項分布の確率であるから，

$$n_{\cdot\cdot}!\prod\prod\frac{(\pi_{ij})^{n_{ij}}}{n_{ij}!}=n_{\cdot\cdot}!\prod\frac{(\pi_{i\cdot})^{n_{i\cdot}}}{n_{i\cdot}!}\times n_{\cdot\cdot}!\prod\frac{(\pi_{\cdot j})^{n_{\cdot j}}}{n_{\cdot j}!}$$
$$\times\frac{\prod n_{i\cdot}!\prod n_{\cdot j}!}{n_{\cdot\cdot}!}\prod\prod\frac{1}{n_{ij}!}\left(\frac{\pi_{ij}}{\pi_{i\cdot}\pi_{\cdot j}}\right)^{n_{ij}}\qquad(6\mathrm{d}.1.1)$$

となる．いま $\pi_{ij}=\pi_{i\cdot}\pi_{\cdot j}$，つまり，属性(事象)$A_i$ と B_j とは独立とすれば，(6d. 1. 1) は

$$n_{\cdot\cdot}!\prod\frac{(\pi_{i\cdot})^{n_{i\cdot}}}{n_{i\cdot}!}\times n_{\cdot\cdot}!\prod\frac{(\pi_{\cdot j})^{n_{\cdot j}}}{n_{\cdot j}!}\times\frac{\prod n_{i\cdot}!\prod n_{\cdot j}!}{n_{\cdot\cdot}!}\prod\prod\frac{1}{n_{ij}!}\qquad(6\mathrm{d}.1.2)$$

となる．この式のはじめの 2 項は周辺和の得られる確率を示し，第 3 項は周辺和を固定したときの各階級の度数が得られる確率を示している．周辺和 $n_{1\cdot},\cdots,n_{r\cdot}$ および $n_{\cdot 1},\cdots,n_{\cdot s}$ を与えたときにすべての n_{ij} が得られる条件つき確率は

$$\frac{\prod n_{i\cdot}!\prod n_{\cdot j}!}{n_{\cdot\cdot}!}\prod\prod\frac{1}{n_{ij}!}\qquad(6\mathrm{d}.1.3)$$

である．式 (6d. 1. 3) は，もちろん，2 つの属性は独立として，仮説の割合 $\pi_{i\cdot}$，$\pi_{\cdot j}$ とは無関係であることに注目すると興味深い．

ある情況，特に計画された実験においては，一方の周辺和が事前に定められている．[たとえば，何人かの人を選んで伝染病の予防接種をするとする．このとき，別に選んだ何人かを対照とすることができるであろう．各グループ（予防接種をした人としなかった人）で，その病気に感染した人としなかった人の実数がわかり，これから 2 × 2 の分割表がつくられる．] 一般に，行和が前もって固定していれば，各行の異なるカテゴリーに属する確率の組は同じ，つまり $\pi_1,\cdots,\pi_s$ と仮定すると，観測度数の得られる同時確率は

$$P(n_{11},\cdots,n_{rs}\mid n_{1\cdot},\cdots,n_{r\cdot})=\prod\frac{n_{i\cdot}!}{n_{i1}!\cdots n_{is}!}\pi_1^{n_{i1}}\cdots\pi_s^{n_{is}}\qquad(6\mathrm{d}.1.4)$$

となる．周辺和 $n_{\cdot j}$ の確率は，この場合

$$P(n_{\cdot 1},\cdots,n_{\cdot s}\mid n_{1\cdot},\cdots,n_{r\cdot})=\frac{n_{\cdot\cdot}!}{n_{\cdot 1}!\cdots n_{\cdot s}!}\pi_1^{n_{\cdot 1}}\cdots\pi_s^{n_{\cdot s}}\qquad(6\mathrm{d}.1.5)$$

であるが，これは $n_{\cdot j}=\sum_i n_{ij}$，$j=1,\cdots,s$ の条件のもとで (6d. 1. 4) を加えて得られる．したがって周辺和を与えたときの n_{ij} の条件つき確率は，(6d. 1. 4)/(6d. 1. 5) であるから

$$\frac{\prod n_{i\cdot}!\prod n_{\cdot j}!}{n_{\cdot\cdot}!}\prod\prod\frac{1}{n_{ij}!}\qquad(6\mathrm{d}.1.6)$$

となる．これは一般の場合に得られた (6d. 1. 3) の式と同じである．

6d. 2　分割表における独立性の検定

分割表における各区画の確率 π_{ij} が指定されると，データがこれらの仮説値と一致するという仮説を検定するには，統計量

$$\chi^2=\sum_i\sum_j\frac{(n_{ij}-n_{\cdot\cdot}\pi_{ij})^2}{n_{\cdot\cdot}\pi_{ij}}=\sum\frac{(O-E)^2}{E}\quad(\text{すべての級について})\qquad(6\mathrm{d}.2.1)$$

が自由度 $(rs-1)$ の χ^2 分布をすることが利用できる．ここで唯一の制約は $\sum\sum n_{ij}=n_{\cdot\cdot}$ である．もし属性が独立ならば，区画の確率は

$$\pi_{ij}=\pi_{i\cdot}\pi_{\cdot j}\qquad(\text{すべての } i \text{ および } j \text{ について})\qquad(6\mathrm{d}.2.2)$$

という関係をみたす．では，この仮説は観測データをもとにしてどのように検定できるか．これには次の 2 つの場合があろう．

1. 周辺分布を規定する仮説的確率 $\pi_{i\cdot}$ と $\pi_{\cdot j}$ がわかっている場合で，このときには区画の確率が (6d. 2. 2) の規則でつくれるかどうかを検討しなければならない．
2. 周辺度数の仮説的比率がわからない場合で，属性が独立であるかどうかの検定をすることが要求される．

まず第 1 の場合には

$$\chi^2=\sum\sum\frac{(n_{ij}-n_{\cdot\cdot}\pi_{i\cdot}\pi_{\cdot j})^2}{n_{\cdot\cdot}\pi_{i\cdot}\pi_{\cdot j}},\qquad\text{自由度}=rs-1\qquad(6\mathrm{d}.2.3)$$

を求めれば，観測度数と期待度数との総合的な食違いの程度が測られる．この式から，観測周辺度数とその期待度数との差を測る 2 つの成分

$$\chi_1^2=\sum_i\frac{(n_{i\cdot}-n_{\cdot\cdot}\pi_{i\cdot})^2}{n_{\cdot\cdot}\pi_{i\cdot}}\qquad\text{自由度}=r-1\qquad(6\mathrm{d}.2.4)$$

$$\chi_2^2=\sum_j\frac{(n_{\cdot j}-n_{\cdot\cdot}\pi_{\cdot j})^2}{n_{\cdot\cdot}\pi_{\cdot j}}\qquad\text{自由度}=s-1\qquad(6\mathrm{d}.2.5)$$

をとり出すことができる．これらの統計量を用いて，観測した周辺和がその期待値と合うかどうかの検定ができる．全体の χ^2 から χ_1^2,χ_2^2 を引いて

$$\chi_3^2=\chi^2-\chi_1^2-\chi_2^2$$

$$= \sum\sum \frac{[n_{ij} - n_{i\cdot}\pi_{\cdot j} - n_{\cdot j}\pi_{i\cdot} + n_{\cdot\cdot}\pi_{i\cdot}\pi_{\cdot j}]^2}{n_{\cdot\cdot}\pi_{i\cdot}\pi_{\cdot j}} \qquad (6d.2.6)$$

を得る．(6d.2.3) の χ^2 は自由度 $rs-1$ であり，χ_1^2 は $r-1$，χ_2^2 は $s-1$ の自由度をもち，かつ χ_3^2 は正であるから，[3b.4 の (iv)] の結果の拡張より，χ_3^2 の自由度は $(rs-1)-(r-1)-(s-1)=(r-1)(s-1)$ となることがわかる．(6d.2.6) の成分 χ_3^2 は独立性からのふれの検定に利用できる．

例として，次の 2×2 分割表で，属性 A については仮説の比率が (p_1, q_1)，属性 B については (p_2, q_2) である場合を考えよう．

	B_1	B_2	計
A_1	a	b	$a+b$
A_2	c	d	$c+d$
計	$a+c$	$b+d$	n

全体の χ^2 は

$$\chi^2 = \frac{(a-np_1p_2)^2}{np_1p_2} + \frac{(b-np_1q_2)^2}{np_1q_2} + \frac{(c-nq_1p_2)^2}{nq_1p_2} + \frac{(d-nq_1q_2)^2}{nq_1q_2} \qquad \text{自由度 } 3$$

である．各成分は次のようになる．

$$\chi_1^2 = \frac{(a+b-np_1)^2}{np_1} + \frac{(c+d-nq_1)^2}{nq_1} = \frac{(a+b-np_1)^2}{np_1q_1} \qquad \text{自由度 } 1$$

$$\chi_2^2 = \frac{(a+c-np_2)^2}{np_2} + \frac{(b+d-nq_2)^2}{nq_2} = \frac{(a+c-np_2)^2}{np_2q_2} \qquad \text{自由度 } 1$$

$$\chi_3^2 = \frac{(aq_1q_2 - bp_2q_1 - cp_1q_2 + dp_1p_2)^2}{np_1p_2q_1q_2} \qquad \text{自由度 } 1$$

例 1 表 6d.2α は，ベーツソンが交雑によって得たスイートピーの個体の分布を示す．周辺度数は 3 : 1 の比になるものと期待され，さらに 2 つの性質，花の色と花粉の形が独立に遺伝するならば，区画の度数は 9 : 3 : 3 : 1 の比となるはずである．

表 6d.2α スイートピーの花粉の形および花の色による個体分布

	花の色		
花粉の形	紫	赤	計
長い	296	27	323
丸い	19	85	104
計	315	112	427

自由度 3 の不一致度を見る全体の χ^2 は 222.1221 で，非常に大きいことはすでに見た (6b.5，例 1)．第 1 の成分

$$\chi_1^2 = \frac{(a+b-np_1)^2}{np_1q_1} = \frac{\left(323 - \frac{3}{4}\times 427\right)^2}{427\times\frac{3}{4}\times\frac{1}{4}} = 0.0945$$

は，自由度 1 で，きわめて小さいから，花粉の形に関する単因子分離は期待どおりであることがわかる．同様に $\chi_2^2 = 0.3443$ もまたきわめて小さいから，花の色についての単因子分離も期待どおりであることがわかる．第 3 の成分は

$$\chi_3^2 = \frac{(aq_1q_2 - bp_2q_1 - cp_1q_2 + dp_1p_2)^2}{np_1p_2q_1q_2}$$

$$= \frac{(296 - 27\times 3 - 19\times 3 + 85\times 9)^2}{427\times 3\times 3\times 1\times 1} = 221.6833$$

となり，自由度 1 できわめて大きい．3 つの χ^2 の値の合計は

$$0.0945 + 0.3443 + 221.6833 = 222.1221$$

となり，先に計算した全体の値と一致する．これから，全体の不一致が自由度 1 の 1 つの成分に集中していることがわかる．つまり，各区画の観測度数の期待度数からのずれは，遺伝した性質の間の従属性によるのであって，単因子分離によるものではないことがわかった．すべての統計的検定が成功するか否かは，問題点を判断するのにより有効な成分を分離できるか否かによる．

つぎに，周辺確率の仮説の値は未知とする．このときには

$$\pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$$

という仮説に基づいて，それらの値を推定する．その最尤推定値は

$$\hat{\pi}_{i\cdot} = \frac{n_{i\cdot}}{n_{\cdot\cdot}} \quad と \quad \hat{\pi}_{\cdot j} = \frac{n_{\cdot j}}{n_{\cdot\cdot}}$$

である．これらの値は全体の χ^2 の計算式に代入され，

$$\chi^2 = \sum\sum \frac{\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n_{\cdot\cdot}}\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n_{\cdot\cdot}}} \qquad (6d.2.7)$$

を得る．$(r-1)+(s-1)$ 個のパラメータ（母数）が推定されたから，6b.2 で導いた適合度の χ^2 分布についての結果を用いると，(6d.2.7) の χ^2 は $(rs-1)-(r-1)-(s-1)=(r-1)(s-1)$ の自由度をもつ．$\pi_{i\cdot}$ と $\pi_{\cdot j}$ の推定値を用いるときは，成分 χ_1^2 と χ_2^2 の値は 0 となり，$\chi_3^2 = \chi^2$ となる．したがって χ_3^2 は独立性からのずれを測ることになる．

この検定は次の 2 つの場合に有用である．

1. ベーツソンの遺伝因子の分離の問題で，周辺度数がその期待度数から著しくずれていることがわかったとする．これは，指定した周辺確率が正しくないことを示す．実際，2 つの型の個体の生活力が異なるために単因子分離が乱されるときには，周辺度数は 3 : 1

の比にならないであろう．そのような場合には，(6d. 2. 6) の第3の成分 χ_3^2 はその正当性を失ってしまう．いいかえると，χ_3^2 の有意性はこの誤った比率の利用から起こるかもしれないのである．この場合，最善のやり方は，比率の推定値を代入し，さきに求めた(6d. 2. 7) の検定を用いることである．

2.　第2の場合は，周辺の比率についてなにも規定されていないときに起こる．この場合には (6d. 2. 7) は独立性の検定を与える．

2×2 の表で仮説の周辺比率が不明のときには，(6d. 2. 7) の独立性の χ^2 検定は

$$\frac{[a-(a+b)(a+c)/n]^2}{(a+b)(a+c)/n}+\frac{[b-(a+b)(b+d)/n]^2}{(a+b)(b+d)/n}+\cdots+\cdots$$

となり，これは

$$n(ad-bc)^2\div(a+c)(b+d)(a+b)(c+d)$$

と整理できる．因子の分離に関するベーツソンの問題では，その数値は

$$\frac{427(296\times85-27\times19)^2}{315\times112\times323\times104}=218.8722$$

となり，自由度は1である．218.8722 という χ^2 の値は，周辺比率に仮説値を用いて得た値よりも小さい．しかし，この程度の差では一般には正反対の結論にはならない．さきの検定では 427 個体の全体によって与えられる情報を利用したが，今度の検定では同じ周辺和をもつような結果の組から得られる情報を用いている．周辺和のあるものは平均よりも情報量が多いが，他のものは少ない．

例 2　表 6d. 2β のデータは，3つの異なる時期にある地方から発掘された沢山の頭蓋骨の性別分布を示している．最初の2つの時期には研究者 A が，第3の時期には研究者 B が男女の性別分類をした．このデータは，その地方に埋められている母集団の性比についてどのような情報を与えるか？

表 6d. 2β　頭蓋骨の性と時期による分布

	時期			計
	1	2	3	
♂	162	180	210	552
♀	110	125	200	435

表 6d. 2γ　性比が等しいという仮説の下での期待度数

	時期		
	1	2	3
♂	136	152.5	205
♀	136	152.5	205

この問題にどのように接近すればよいであろうか．まず，各時期において男女比は 1 : 1 といえるかどうかを調べるのが自然である．そのような仮説のもとでの期待値は表 6d. 2γ に示されている．第1時期の χ^2 は

$$\frac{(162-136)^2}{136}+\frac{(110-136)^2}{136}=9.9412$$

となり，第2，第3時期についてはそれぞれ 9.9180，0.2440 となり，χ^2 の合計は 20.1032 で自由度は3となる．この χ^2 の値はきわめて大きく，性比は 1 : 1 から離れていることがわかる．個々に見ると，はじめの2時期での偏差が有意である（χ^2 はそれぞれ自由度1で 9.9412 と 9.9180 である）．1 : 1 の性比からの偏差を周辺和から測った χ^2 は

$$\frac{(552)^2}{493.5}+\frac{(435)^2}{493.5}-987=13.8692$$

であるから，自由度2の $\chi^2=20.1032-13.8692=6.2340$ が残る．これは3つの時期で性比が異なるという仮説を検定するためのものである．この値は疑いもなく5％水準で有意であって，性比に差があることを示すが，この検定は厳密には正しくない．なぜなら $\chi^2=13.8692$，自由度1は大きく，周辺和は性比 1 : 1 と相いれないという事実があるからである．すでに見たように，1 : 1という性比から全体としてずれているのであるから，3時期の性比は 1 : 1 ではなくても同じであるかどうかを調べてみよう．周辺和を固定したときの独立性の検定 (6d. 2. 7)，または類似標本の一様性の検定 (6c. 2. 2) をここではそのまま計算することができる．これによると χ^2 は 6.3222 で自由度は2で，3時期の性比に有意な差があることがわかる．この χ^2 がさきに仮説の比を用いて求めた値 6.2340 とよく一致したのは，多分偶然によるものであろう．

この種のデータから結論を導くときには，われわれは注意深くなければならない．頭蓋骨の性別は解剖学的鑑識という主観的な方法で行われるから，鑑定者が異なるとその性別のしかたは異なり，したがってその性比も変わってくるかもしれないということに気がつくであろう．だから異なる時期の性比に見られる食違いは，第3時期の鑑定者が別の人であったことによるのかもしれないのである．観測結果の男の比率は次のようになる．

	研究者A		研究者B	
時　期	1	2	3	全体
男の比率	0.5956	0.5902	0.5122	0.5593

このとき，この2人の研究者の差は，次の 2×2 分割表の独立性の χ^2 検定によって検定できる．

	第1，第2時期	第3時期
♂	342	210
♀	235	200

この χ^2 の値は 6.3055 で，自由度1で有意である．したがって，性比の差を測る自由度2の全体の χ^2 の値 6.3222 のうち，1自由度だけで 6.3055 となるものがあり，これは，時期間の差の全体が研究者の違いだけによることを示している．

さて結論はどうなるのであろうか？　研究者たちはその地方の同じ領域で頭蓋骨の性鑑定を行っていたとしても，かれらの間にその技術に差があるのであるから，2人の研究者の判定結果は比較できないのである．

このデータはまた，男性の比率が優勢であることを示している．1：1からのこのずれがどの程度まで頭蓋骨の性別鑑定の誤った技術によるのかは不明である．また男性の頭蓋骨と女性の頭蓋骨の保存性に差があるかもしれないという問題もある．事実，女性の頭蓋骨は男性のものに比べて繊細でもろいのである．このようにして，"1つの問題を解決するのに，10以上の問題をつくり出す"ことになるようである．

例 3　いくつかの学校から集めた，児童の言語能力の欠陥 (S_1, S_2, S_3) と身体上の欠陥 (P_1, P_2, P_3) に関する次のデータを考えよう．

	S_1	S_2	S_3	計
P_1	45	26	12	83
P_2	32	50	21	103
P_3	4	10	17	31
計	81	86	50	217

独立性の仮説の下での期待値は次のようになる：

30.982	32.894	19.124
38.447	40.820	23.733
11.571	12.286	7.143

これから自由度 $2 \times 2 = 4$ の χ^2 を求めると 32.8828 となり，1％水準で有意となる．1つの区画での度数は4と小さいのに，その期待度数は十分に大きくて，χ^2 近似が成り立つことがわかる．しかし近似が良いためには，観測度数も大分大きくなければならない．このような場合には，2つの区画を一緒にし，その度数を加えそれらを1つの区画として取り扱って有意性の検定をするのが合理的である．さきの例では，4と10の区画は1つにして度数を14とすれば対応する期待値は $11.571 + 12.286 = 23.857$ となる．この8個の区画について公式 $\sum(0 - E)^2/E$ の式で新しく計算した χ^2 は 31.5762 となる．ただし，この値の分布は，もはや χ^2 分布ではないが，実用上は自由度が1少ない，すなわち $4 - 1 = 3$ の自由度の χ^2 分布として引用する．和をとるときは1区画分だけ少なくなるが，理論的に1自由度が失なわれるのではない．それゆえ新しい χ^2 を自由度 $4 - 1 = 3$ として扱うと有意性を過大評価し，一方自由度4と考えると有意性を過小評価することになる．新しい χ^2 が自由度3で有意ではないか，または，この例の場合のように自由度4で有意のときは一定の結論が出せるが，新しい χ^2 が自由度3で有意であり，自由度4で有意で

ないというときには，正しい有意水準について推測することはできない．

このような場合の χ^2 の適切な分布は Chernoff and Lehmann (1954) によって研究されている．かれらによると，この分布は χ^2 変量の1次結合の分布となる．しかし，そのような分布を用いるのは難かしいので，次の方法が示唆されている．まず，2つの区画 S_1P_3 と S_2P_3 は1つの区画をつくっているものと考えて，新しい組の期待値を計算する．いま p_1, q_1, r_1 と p_2, q_2, r_2 とを身体上の欠陥と言語能力の欠陥の周辺比率とすると，これらが独立という仮説のもとで観測度数が得られる確率は

$$(p_1p_2)^{45}(p_1q_2)^{26}(p_1r_2)^{12}\cdots[(p_2 + q_2)r_1]^{14}(r_1r_2)^{17}$$

に比例する．p_1, q_1, p_2, q_2 を求める最尤方程式は

$$217p_1 = 83 \qquad 217q_1 = 103$$
$$76p_2 = 77q_2 \qquad 217(1 - p_2 - q_2) = 50$$

となるから，

$$\hat{p}_2 = 0.387308, \quad \hat{q}_2 = 0.382278, \quad \hat{r}_2 = 0.230414$$

という推定値が求まる．p_1, q_1, r_1 の推定値は前と変わらない．新しい期待度数は

32.147	31.729	19.124
39.893	39.375	23.733
(12.006 +	11.851)	7.143

となり，自由度 $7 - 4$ の χ^2 の値は 31.2472 となる．この検定は，有意性が認められるとき独立性の仮説が棄却されるという意味で正当である．

フィッシャーは，区画の度数が小さいときにより適切な，尤度に基づく検定（6e 参照）を推奨した．これは

$$L = 2\sum\sum n_{ij}\log_e n_{ij} - \sum n_{i.}\log_e n_{i.} - \sum n_{.j}\log_e n_{.j} + n_{..}\log_e n_{..}$$

で定義され，L は近似的に自由度 $(r-1)(s-1)$ の χ^2 分布に従う．さきの問題での L の値は 30.4448 であり，これは自由度4で有意である．この L を用いる場合でも大標本は必要である．

検定の目的は，独立性からの一般的なずれをまず確認することである点を強調しておく必要がある．このためには，計算の簡単な，根拠の確実な検定を用いれば十分である．そして必要に応じて，分割表のいろいろの部分を吟味するための，より洗練された検定を用いればよい．たとえば，2種の身体上の欠陥 P_2, P_3 と言語能力の欠陥 S_1, S_2 が関連しているかどうかを調べることができる．これには，次節で述べる洗練された手法が必要である．

6d.3　小標本における独立性の検定

母数が満足すべきある関係を規定する任意の仮説の検定においては，未知母数が χ^2 統

計量の大標本分布にふくまれないことを知った．しかし，小標本においては χ^2 近似が成り立たなくなったり，さらには χ^2 の正確な分布に未知母数が現れたりすることが起こる．この後の場合には，攪乱母数と呼ばれる未知母数が確率分布のなかに含まれるために，正確な有意性検定は不可能になる．

攪乱母数を除去する1つの方法は，観測された特定の標本を，攪乱母数の正確な値がもしわかっているとしたらそれとの比較ができるであろうような標本の全集団と比較するのではなく，標本の部分集合を選び，そのなかでの統計量の分布はどの未知母数をも含まないようにして，その部分集合と比較することである．たとえば，2つの属性が独立であるという仮説のもとでは，周辺和を与えたときに区画の度数が得られる確率は

$$\frac{\prod n_{i.}!\prod n_{.j}!}{n_{..}!}\prod\prod\frac{1}{n_{ij}!}$$

となり，これは仮説の値 $\pi_{i.}$ と $\pi_{.j}$ を含まない．したがって，独立性の検定において正確な確率を求めうる可能性があるのである．

例 1　都市 (C) から集めた兵士と農村 (V) から集めた兵士の社交性 (S) と非社交性 (NS) に関する次のデータから，都市出身の兵士のほうが農村出身の兵士よりもより社交的であるといえるか？

	S	NS	計
C	13	4	17
V	6	14	20
計	19	18	37

一方の対角要素の度数が比較的小さいことは，都市出身の兵士のほうがより社交的であることを示唆している．しかし，ここで観測された数値やそれ以上に都市出身兵士の社交性を示す数値の全体が，都市出身と農村出身の兵士の間に社交性についてほんとうになんの差もなかったとした場合に，偶然に起こることはないかどうかを確かめておかねばならない．周辺和を固定したとき，4区画における観測度数が所与の a, b, c, d である確率は

$$\frac{17!\ 20!\ 19!\ 18!}{37!}\frac{1}{a!b!c!d!}$$

であるから，右上の区画の数値が 4, 3, 2, 1, 0（これらの数値は独立性の仮説からより遠ざかり，示唆された対立仮説を支持する方向にある）である．確率は，それぞれ 0.0^25218, 0.0^35966, 0.0^43607, 0.0^51097, 0.0^71075 となり，その合計は 0.0059 となる．この確率は非常に小さいから，都市出身の兵士が農村出身の兵士よりもより社交的であることが指摘される．

区画の度数が小さくなければ，独立性を検定するための χ^2 を計算し，平均0，分散1の正規偏差量が χ を越える確率を求めることによって，この結果を確認することもできるで

あろう．前の例では，

$$\chi^2=\frac{37(13\times14-6\times4)^2}{17\times20\times19\times18}=7.9435,\qquad \chi=\sqrt{7.9435}=2.8181$$

となるから，正規分布の確率は 0.0025 となる．この値は，正しい値 0.0059 よりも小さいが，この食違いは標本の小さいことによる．

Yates (1934) は，周辺和はそのままにして小さいほうの度数*) を 1/2 だけふやして得られる分割表を用いて χ^2 を計算すると，正しい確率により近い近似が得られることを示唆した．この例では，連続性の補正をしたといわれる，この新しい χ^2 は

$$\chi^2=\frac{37(12.5\times13.5-6.5\times4.5)^2}{17\times20\times19\times18}=6.1922$$

となる．この χ の値は 2.4884 であるから，正規分布表から確率は 0.0064 となり，補正しない χ^2 の場合よりも正しい値に近くなる．

これとは少し違う方法が V. M. ダンデカーによって提案されているが，この方法では $\chi_0^2, \chi_{-1}^2, \chi_1^2$ を計算する．これらはそれぞれ観測度数そのもの，もっとも小さい度数を1ふやしたもの，および1減らしたものから計算される．これらから，χ^2 の補正値は

$$\chi_c^2=\chi_0^2-\frac{\chi_0^2-\chi_{-1}^2}{\chi_1^2-\chi_{-1}^2}(\chi_1^2-\chi_0^2)$$

によって求められる．この例では $\chi_0^2=7.9435$, $\chi_1^2=12.0995$, $\chi_{-1}^2=4.6587$ となるから

$$\chi_c^2=7.9435-\frac{7.9435-4.6587}{12.0995-4.6587}(12.0995-7.9435)=6.1086$$

$$\chi_c=2.4715$$

である．したがって，正規分布から確率を求めると 0.0068 でこれも正しい値に近い．

この場合に尤度比検定

$$L=2\sum O\log_e\frac{E}{O}$$

を用いると値 8.2811 を得る．χ の値は $\sqrt{8.2811}=2.8778$ となるから，確率は正しい値よりもずっと小さい．だから，尤度比検定は事態の改良にはならない．

例 2　例1での調査の目的は，都市出身の兵士が農村出身の兵士よりも社交的であるかどうかを調べることであった．このために，期待値からの偏差を1方向でのみ考えた．しかし，一般には，2つの属性間の関連を，その性質を規定せずに発見することが目的であるならば，期待値からの偏差を両方向に考える必要がある．観測された χ^2 よりも大きな χ^2 値を与える分割表は，右上の区画の度数が 3, 2, 1, 0 となる場合だけではなく，左上の区画の度数が 4, 3, 2, 1, 0 となる場合もある．後者の場合それぞれの確率は 0.0^22088, 0.0^31864, 0.0^58733, 0.0^61828, 0.0^81132 となり，合計すると 0.0023 である．右上の

*) この例で小さいほうの度数とは，(6, 4) という対角線上の度数で，農村出身の兵士の社交性と都市出身の兵士の非社交性とを指す．一般の独立性の検定では，χ^2 の値が小さくなるように度数に 1/2 を加える．

区画の度数が小さい度数をとる確率は前に計算したように 0.0059 であるから，これを合計すると全体の確率は 0.0059 + 0.0023 = 0.0082 となり，やはり小さな値で独立性からのずれを示している.

標本が大きいときには，補正前の χ^2 を直接使って自由度1の χ^2 表からその確率を求めることができる．この例では χ^2 は 7.9435 で，それに対応する確率は約 0.005 で，正しい値 0.0082 よりも小さい.

イェーツの連続性補正を行うと χ^2 は 6.1922 となり，対応する確率は 0. 128 で，正しい値よりも大きくなる.

ダンデカーの補正をこの例に施すためには，観測値から求めた $\chi^2 = 7.9435$ よりも小さい χ^2 値と大きい χ^2 値を，分割表

$$\begin{matrix} 12 & 5 \\ 6 & 14 \end{matrix} \quad と \quad \begin{matrix} 13 & 4 \\ 5 & 15 \end{matrix}$$

から求めて，それらが 6.0598 と 9.7448 となることにまず注意する．補正済 χ^2 は

$$\chi_c^2 = 7.9435 - \frac{7.9435 - 6.0598}{9.7448 - 6.0598}(9.7448 - 7.9435)$$
$$= 7.0228$$

となり，確率は 0.0082 で，正しい値にほとんど正確に等しい．一般に，ダンデカーの補正のほうがイェーツの方法よりも少し良いのであるが，その補正の方法はイェーツのほうが簡単である.

2つの遺伝子によって分類されたデータを基礎にして連鎖の有無を検定するとき，組換価が 1/2 よりも小さいことがわかっていれば，一方向の結合関係だけを検定すれば十分である．ところが，フィッシャーによって示されたように，組換価は 1/2 を越えることがありうるのである．それゆえ，結合の性質の調査には立ち入らずに，まず独立性の仮説の誤っていることを示すのがよい．さらに，独立性からのずれは実験データではいろいろの原因によって起こりうることに注意しなければならない．したがって，独立性の仮説との両立性に関して，データを直接評価するような検定を用いるのがよい.

分類値（カテゴリカルデータ）における χ^2 検定の使い方を示したその他の例については，Fisher (1925) を参照されたい.

6e いくつかの一般的な大標本検定

6e.1 記号と基本的な結果

いま $x_1, \cdots, x_n$ を互いに独立で同一分布に従う確率変数で多次元のものであってもよい

とし，その共通の確率密度関数を $p(x, \theta)$ とする．これは q 次元の母数 $\theta' = (\theta_1, \cdots, \theta_q)$ に依存する．θ の許容できる集合は，q 次元ユークリッド空間の退化しない区間とする．n 個の独立な観測値の対数尤度と，**5f, 5g** で定義した第 i 有効スコアはそれぞれ

$$l(\theta) = \log p(x_1, \theta) + \cdots + \log p(x_n, \theta)$$

$$\phi_i(\theta) = \frac{1}{\sqrt{n}} \frac{\partial l}{\partial \theta_i}$$
$$= \frac{1}{\sqrt{n}} \left\{ \frac{1}{p(x_1, \theta)} \frac{\partial p(x_1, \theta)}{\partial \theta_i} + \cdots + \frac{1}{p(x_n, \theta)} \frac{\partial p(x_n, \theta)}{\partial \theta_i} \right\} \qquad (6e.1.1)$$

である．いま $\mathscr{I}$ を1個の観測値 x による θ についての情報行列とする．すなわち，$\mathscr{I} = (\mathscr{I}_{rs})$ とすると，

$$\mathscr{I}_{rs} = E\left\{ \frac{1}{p(x, \theta)} \frac{\partial p}{\partial \theta_r} \cdot \frac{1}{p(x, \theta)} \frac{\partial p}{\partial \theta_s} \right\} \qquad (6e.1.2)$$

である.

（i） いま $\mathbf{V}' = (\phi_1(\theta), \cdots, \phi_q(\theta))$ とする．このとき $\mathbf{V}$ の漸近分布は，平均 0，分散共分散行列 $\mathscr{I}$ の q 変量正規分布である.

これを証明するには，線型関数 $\mathbf{b}'\mathbf{V} = b_1\phi_1(\theta) + \cdots + b_q\phi_q(\theta)$ の漸近分布が，平均 0，分散が $\mathbf{b}'\mathscr{I}\mathbf{b}$ の1変量正規分布であることを示せば十分である．いま

$$y_i = \frac{b_1}{p(x_i, \theta)} \frac{\partial p(x_i, \theta)}{\partial \theta_1} + \cdots + \frac{b_q}{p(x_i, \theta)} \frac{\partial p(x_i, \theta)}{\partial \theta_q} \qquad (6e.1.3)$$

を考える．このとき，$E(y_i) = 0$ であり

$$V(y_i) = \sum \sum b_r b_s E\left\{ \frac{1}{p(x_i, \theta)} \frac{\partial p}{\partial \theta_r} \cdot \frac{1}{p(x_i, \theta)} \frac{\partial p}{\partial \theta_s} \right\}$$
$$= \sum \sum b_r b_s \mathscr{I}_{rs} = \mathbf{b}'\mathscr{I}\mathbf{b}$$

である．それゆえ $y_1, \cdots, y_n$ はいずれも平均 0，分散 $\mathbf{b}'\mathscr{I}\mathbf{b}$ をもち，互いに独立に同一分布に従う確率変数である．よって $n \to \infty$ のときの $(y_1 + \cdots + y_n)/\sqrt{n}$ の漸近分布は $N(0, \mathbf{b}'\mathscr{I}\mathbf{b})$ となる．ところで

$$(y_1 + \cdots + y_n)/\sqrt{n} = \mathbf{b}'\mathbf{V} \qquad (6e.1.4)$$

である．よって (i) が証明された.

（ii） $p(x, \theta)$ が **5f.1** の正則条件をみたすものとする．$\hat{\theta}$ を n 個の観測値に基づく尤度方程式の根の一致推定量とし $\mathbf{D}'$ は偏差ベクトル $[\sqrt{n}(\hat{\theta}_1 - \theta_1), \cdots, \sqrt{n}(\hat{\theta}_q - \theta_q)]$ とする．さらに情報行列 $\mathscr{I}$ は非特異とする．このとき次の漸近同等性が成り立つ.

$$\mathbf{V} \overset{a}{=} \mathscr{I}\mathbf{D} \quad つまり \quad \mathbf{D} \overset{a}{=} \mathscr{I}^{-1}\mathbf{V} \qquad (6e.1.5)$$

$$2[l(\hat{\theta}) - l(\theta)] \overset{a}{=} \mathbf{D}'\mathscr{I}\mathbf{D}. \qquad (6e.1.6)$$

(6e.1.5) と (6e.1.6) の2つの結果は，**5f.2** の議論をたどれば簡単に証明される．ほ

とんどの例では最尤推定量は尤度方程式の根の一致推定量となっているので，本節の以下の議論ではそれらは同一と仮定する．

（iii）　q 個の母数 $\theta_1, \cdots, \theta_q$ は $k < q$ 個の制約条件 $R_1(\theta) = 0, \cdots, R_k(\theta) = 0$ をみたすとし，$R_1, \cdots, R_k$ は連続な1階偏導関数をもつとする．この条件は，各 θ_i を $s = q - k$ 個の新しい母数 $\beta_1, \cdots, \beta_s$ の関数

$$\theta_i = g_i(\beta_1, \cdots, \beta_s)$$

として規定するのと同等であるとする．ここで g_i は連続な1階偏導関数をもつとする．$\hat{\beta}$ を β の最尤推定量とし，さらに

$$\Psi_i(\beta) = \partial l / \sqrt{n}\, \partial \beta_i, \qquad \mathbf{U}' = [\Psi_1(\beta), \cdots, \Psi_s(\beta)]$$

$$\mathbf{F}' = [\sqrt{n}(\hat{\beta}_1 - \beta_1), \cdots, \sqrt{n}(\hat{\beta}_s - \beta_s)]$$

とする．また $\mathbf{H}$ を1個の観測値 x による β に関する情報行列とする．このとき

（a）　$\mathbf{U} \overset{a}{=} \mathbf{HF}$　つまり　$\mathbf{H}^{-1}\mathbf{U} \overset{a}{=} \mathbf{F}$　(6e. 1. 5) より　(6e. 1. 7)

（b）　$2[l(\hat{\beta}) - l(\beta)] \overset{a}{=} \mathbf{F}'\mathbf{HF}$　(6e. 1. 8)

（c）　$\mathbf{U} = \mathbf{M}'\mathbf{V}$　かつ　$\mathbf{H} = \mathbf{M}'\mathscr{I}\mathbf{M}$　(6e. 1. 9)

を得る．ここに，$\mathbf{M} = (\partial g_i / \partial \beta_j)$ は次数 $q \times s$ の行列である．

この (b) は (6e. 1. 6) と同様であり，(c) を証明するのは容易である．

6e. 2　単純仮説の検定

6e. 1 の条件のもとで $\theta = \theta_0$（指定した値）という仮説を検定するために，Neyman and Pearson (1928) は尤度比基準

$$\Lambda_0 = \frac{L(\theta_0)}{\sup_{\theta} L(\theta)} \tag{6e. 2. 1}$$

を提案した．ここに sup は θ の許容できる集合全体にわたってとるものとする．対数をとって θ の最尤推定量 $\hat{\theta}$ を使うと，検定基準 (6e. 2. 1) は

$$-2 \log_e \Lambda_0 = 2[l(\hat{\theta}) - l(\theta_0)] \tag{6e. 2. 2}$$

と書ける．

$\mathbf{D}_0, \mathbf{V}_0, \mathscr{I}_0$ を θ のかわりに θ_0 をおいたときの $\mathbf{D}, \mathbf{V}, \mathscr{I}$ の値とし，$\mathscr{I}(\hat{\theta})$ を $\theta = \hat{\theta}$ での $\mathscr{I}$ の値とする．仮説 $\theta = \theta_0$ を検定するのに，Wald (1943) は基準

$$W_0 = \mathbf{D}_0'[\mathscr{I}(\hat{\theta})]\mathbf{D}_0 \tag{6e. 2. 3}$$

を提案し，著者 (Rao (1947e)) は有効スコアに基づく基準

$$S_0 = \mathbf{V}_0'\mathscr{I}_0^{-1}\mathbf{V}_0 \tag{6e. 2. 4}$$

を提案した．後者では最尤推定値をはっきりした形で計算する必要がない．(6e. 1. 6)から

$$2[l(\hat{\theta}) - l(\theta_0)] \overset{a}{=} \mathbf{D}_0'\mathscr{I}_0\mathbf{D}_0 \overset{a}{=} \mathbf{D}_0'[\mathscr{I}(\hat{\theta})]\mathbf{D}_0$$

がわかる．また (6e. 1. 5) から

$$\mathbf{D}_0'\mathscr{I}_0\mathbf{D}_0 \overset{a}{=} \mathbf{V}_0'\mathscr{I}_0^{-1}\mathbf{V}_0$$

であるから，帰無仮説が正しいとき，3つの統計量 $-2\log_e \Lambda_0, W_0, S_0$ は大標本では同等となる．[(i), **6e. 1**] から，$\mathbf{V}$ の漸近分布は平均0，分散共分散行列 $\mathscr{I}$ の q 変量正規分布となる．したがって，$\mathbf{V}_0'\mathscr{I}_0^{-1}\mathbf{V}_0$ の漸近分布は $\chi^2(q)$ となり，3つの統計量はすべて $\mathbf{V}_0'\mathscr{I}_0^{-1}\mathbf{V}_0$ と同等となるから，それらは同じ漸近分布をもつ．

いままでのところ統計学の文献には，帰無仮説からのずれを検出しようとする，これらの検定の相対的な利点についての適当な議論は見当らない．

6e. 3　複合仮説の検定

複合仮説として，$\theta_i, i = 1, \cdots, q$ は $\beta_1, \cdots, \beta_s$ の関数

$$\theta_i = g_i(\beta_1, \cdots, \beta_s) \tag{6e. 3. 1}$$

とする．あるいは，θ_i が $k = q - s$ 個の制約条件

$$R_i(\theta) = 0, \quad i = 1, \cdots, k \tag{6e. 3. 2}$$

をみたすものとする．ここで，g_i と R_i とは連続な1階偏導関数をもつ関数とする．$\hat{\theta}$ を制約条件をつけないときの θ の最尤推定量，θ^* を制約条件をつけたときの最尤推定量とする．これは β の最尤推定量 $\hat{\beta}$ によって

$$\theta_i^* = g_i(\hat{\beta}_1, \cdots, \hat{\beta}_s) \tag{6e. 3. 3}$$

と書ける．

複合仮説にたいするネイマン＝ピアソンの尤度比検定は

$$-2\log_e \Lambda_c = 2[l(\hat{\theta}) - l(\theta^*)] = 2[l(\hat{\theta}) - l(\hat{\beta})] \tag{6e. 3. 4}$$

である．ここに $l(\hat{\beta})$ は対数尤度を β の関数と考えて計算される．このことから

$$\Lambda_c = \frac{\sup_{R_i(\theta)=0} L(\theta)}{\sup_{\theta} L(\theta)} \tag{6e. 3. 5}$$

であることがわかる．ここで，分子の上限は θ が制約条件 $R_i(\theta) = 0,\ i = 1, \cdots, k$ に従うときに，分母の上限は θ のすべての許容値にわたって求める．

複合仮説にたいするラオの有効スコア検定は

$$S_c = \mathbf{V}^{*\prime}\mathscr{I}^{*-1}\mathbf{V}^*, \qquad \text{ただし} \qquad \mathbf{V}^* = \mathbf{V}(\theta^*) \text{ など} \tag{6e. 3. 6}$$

で，また，ワルドの検定は

$$W_c = \sum\sum \lambda^{ij}(\hat{\theta}) R_i(\hat{\theta}) R_j(\hat{\theta}) \tag{6e. 3. 7}$$

で，これは条件をつけないときの θ の最尤推定量だけを用いる．(6e. 3. 7) の式で (λ^{ij}) は，$\sqrt{n} R_i(\hat{\theta})$ の漸近分散共分散行列 (λ_{ij}) の逆行列で，

$$(\lambda_{ij}) = \mathbf{T}'\mathscr{I}^{-1}\mathbf{T}, \qquad \mathbf{T} = \left(\frac{\partial R_i(\theta)}{\partial \theta_j}\right) \quad q \times k \text{ 行列} \tag{6e.3.8}$$

で定義される．(6e.3.5) から (6e.3.7) までの3つの統計量はすべての同じ漸近分布 $\chi^2(q-s)$ をもつ．(6e.1.6) から

$$2[l(\hat{\theta}) - l(\theta)] \stackrel{a}{=} \mathbf{D}'\mathscr{I}\mathbf{D} \stackrel{a}{=} \mathbf{V}'\mathscr{I}^{-1}\mathbf{V}$$

が得られ，(6e.1.7) から (6e.1.9) までの式から

$$2[l(\hat{\beta}) - l(\beta)] \stackrel{a}{=} \mathbf{F}'\mathbf{H}\mathbf{F} \stackrel{a}{=} \mathbf{U}'\mathbf{H}^{-1}\mathbf{U} \stackrel{a}{=} \mathbf{V}'\mathbf{M}\mathbf{H}^{-1}\mathbf{M}'\mathbf{V}$$

がわかる．したがって

$$2[l(\hat{\theta}) - l(\hat{\beta})] \stackrel{a}{=} \mathbf{V}'(\mathscr{I}^{-1} - \mathbf{M}\mathbf{H}^{-1}\mathbf{M}')\mathbf{V} \tag{6e.3.9}$$

となる．(6e.3.9) の右辺の漸近分布が χ^2 分布となる必要十分条件は

$$(\mathscr{I}^{-1} - \mathbf{M}\mathbf{H}^{-1}\mathbf{M}')\mathscr{I}(\mathscr{I}^{-1} - \mathbf{M}\mathbf{H}^{-1}\mathbf{M}') = (\mathscr{I}^{-1} - \mathbf{M}\mathbf{H}^{-1}\mathbf{M}') \tag{6e.3.10}$$

である．(6e.3.10) の左辺は，(6e.1.9) を用いて

$$\begin{aligned}
&\mathscr{I}^{-1}\mathscr{I}\mathscr{I}^{-1} - \mathscr{I}^{-1}\mathscr{I}\mathbf{M}\mathbf{H}^{-1}\mathbf{M}' - \mathbf{M}\mathbf{H}^{-1}\mathbf{M}'\mathscr{I}\mathscr{I}^{-1} + \mathbf{M}\mathbf{H}^{-1}(\mathbf{M}'\mathscr{I}\mathbf{M})\mathbf{H}^{-1}\mathbf{M} \\
&\quad = \mathscr{I}^{-1} - 2\mathbf{M}\mathbf{H}^{-1}\mathbf{M}' + \mathbf{M}\mathbf{H}^{-1}\mathbf{H}\mathbf{H}^{-1}\mathbf{M}' = \mathscr{I}^{-1} - \mathbf{M}\mathbf{H}^{-1}\mathbf{M}'
\end{aligned}$$

となるから，条件 (6e.3.10) はみたされる．χ^2 の自由度は

$$\begin{aligned}
\operatorname{trace}(\mathscr{I}^{-1} - \mathbf{M}\mathbf{H}^{-1}\mathbf{M}')\mathscr{I} &= \operatorname{trace}(\mathbf{I} - \mathbf{M}\mathbf{H}^{-1}\mathbf{M}'\mathscr{I}) \\
&= q - \operatorname{trace}\mathbf{M}\mathbf{H}^{-1}\mathbf{M}'\mathscr{I} \\
&= q - \operatorname{trace}\mathbf{H}^{-1}\mathbf{M}'\mathscr{I}\mathbf{M} \\
&= q - \operatorname{trace}\mathbf{H}^{-1}\mathbf{H} \\
&= q - s
\end{aligned}$$

である．これでネイマン＝ピアソン統計量 (6e.3.4) にたいして求めるべき結果が証明された．統計量 S_c と W_c の漸近分布も同様に証明することができる．

実際には，**6e** で考えたよりももっと一般的な結果を必要とする．たとえば，観測値 $x_1, \cdots, x_n$ が同一の分布に従わないということもあろう．このような場合にも，適当な尤度関数や有効点数を考えてまったく同様に検定基準 Λ, W, S を定義することができる．漸近分布を導くには確率密度についてさらに条件を付加する必要があるが，これは，同一分布には従わない確率変数に多変量中心極限定理を適用可能にすることを保証するためであり，また最尤推定量の望ましい性質 (6e.1.5) を保証するためである．

6f　順序統計量

6f.1　経験分布関数

$x_1, \cdots, x_n$ を，分布関数 $F(x)$ をもつ確率変数 X の互いに独立な n 個の観測値からなる

標本とする．$x_{(i)}$ を，(値が $x_{(i)}$ よりも小さい観測値の個数) $\leqslant (i-1)$ で，(値が $x_{(i)}$ より大きい観測値の個数) $\leqslant (n-i)$ であるような観測値とする．このとき，観測値を大きさの順に，同じ大きさのものがあっても，異なる番号をつけて並べることができるから，

$$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)} \tag{6f.1.1}$$

とできる．ただし，$x_{(1)}$ は観測値の最小値，$x_{(n)}$ は最大値である．

$X_{(i)}(x_1, \cdots, x_n) = x_{(i)}$ となるような，標本値の関数 $X_{(i)}$ を，i 番目の順序統計量とよぶ．

ここで，次のような関数 $S_n(\cdot)$ あるいはもっと正確に書くと $S_n(\cdot \mid x_1, \cdots, x_n)$ を定義しよう．

$$S_n(x) = \begin{cases} 0, & x \leqslant x_{(1)} \text{ のとき} \\ \dfrac{k}{n}, & x_{(r-1)} < x_{(r)} = x_{(r+1)} = \cdots = x_{(k)} < x_{(k+1)} \text{ なら，} x_{(r)} < x \leqslant x_{(k+1)} \\ & \text{のとき} \\ 1, & x_{(n)} < x \text{ のとき} \end{cases}$$

関数 $S_n(\cdot)$ は，観測値に依存するので彷徨関数であり，経験分布関数と呼ばれる．容易にわかるように，$S_n(x)$ は非減少，左連続で，$S_n(-\infty) = 0$，$S_n(\infty) = 1$ であるから，厳密な意味で分布関数である．これは n 個の点で不連続な階段関数である．任意の点 x にたいして，$S_n(x)$ は x より小さい観測値の標本中での割合を示すにすぎず，したがって $0/n, 1/n, \cdots, 1$ という値をとる確率変数である．一方，観測値が x より小さい確率の仮説値は $F(x)$，つまり分布関数の x における値である．したがって，$S_n(x) = m/n$ となる確率は，成功の確率が $F(x)$ である試行を n 回行ったとき，そのうち m 回成功する2項確率，すなわち

$$P\left[S_n(x) = \frac{m}{n}\right] = \binom{n}{m}[F(x)]^m[1 - F(x)]^{n-m} \tag{6f.1.2}$$

と同じである．(6f.1.2) の帰結として，大数の強法則を用いると

$$P\{\lim_{n\to\infty} S_n(x) = F(x)\} = 1, \qquad \text{固定した } x \text{ について} \tag{6f.1.3}$$

あるいは，大数の弱法則を用ると，

$$S_n(x) \xrightarrow{P} F(x), \qquad \text{固定した } x \text{ について} \tag{6f.1.4}$$

が得られる．

$S_n(x)$ の関数の極限でのふるまいについての諸結果を以下に証明なしに述べる．

(i) Glivenko (1933) の定理　$n \to \infty$ のとき x $(-\infty < x < \infty)$ について一様に，系列 $S_n(x) \to F(x)$ となる確率は1である．すなわち記号で書けば $D_n(x) = \sup_x |S_n(x) - F(x)|$ とすると，次のようになる．

$$P(\lim_{n\to\infty} D_n = 0) = 1 \tag{6f.1.5}$$

この結果 (6f. 1. 5) は (6f. 1. 3), (6f. 1. 4) よりも強力である．

(ii)　KOLMOGOROV (1933) の定理　D_n は (i) で定義したとおりで，さらに

$$Q_n(\lambda) = \begin{cases} P(\sqrt{n} D_n < \lambda), & \lambda > 0 \\ 0 & \lambda \leqslant 0 \end{cases}$$

とする．$F(x)$ が連続ならば次のようになる．

$$\lim_{n\to\infty} Q_n(\lambda) = Q(\lambda) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2\lambda^2) & \lambda > 0 \\ 0 & \lambda \leqslant 0 \end{cases}$$

極限分布 $Q(\lambda)$ は，$F(x)$ とはそれが連続であることを除いて無関係であることを確かめよ．これは本章で論じた大標本基準すべての特徴である．この命題の証明については Doob (1949) を見よ．

統計量 $\sqrt{n} D_n$ は与えられた標本が指定された分布からのものであるという仮説の検定に利用できるといわれている．極限分布から求めた $\sqrt{n} D_n$ のパーセント点のいくつかは次のとおりである (Smirnov (1939))．

	0.5%	1%	2.5%	5%	10%
下　側	0.42	0.44	0.48	0.52	0.57
上　側	1.73	1.63	1.48	1.36	1.23

$\sqrt{n} D_n$ に基づくコルモゴロフ＝スミルノフ検定の代替としては，**6b. 1** で述べた χ^2 検定がある．しかし，この χ^2 検定の場合には，x のある選ばれた級区間における観測値の度数分布を必要とする．このとき，観測度数 (O) は，(指定された分布から計算される) 仮説の度数つまり期待度数 (E) と，**6b. 1** で定義した基準

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

を用いて比較される．コルモゴロフ＝スミルノフ検定はある点では鋭敏であるけれども，観測された分布と仮説上の分布とを x の全領域にわたって比較するわけではない．個々の級区間における偏差の検討を通じて求められる χ^2 検定のほうが，指定された分布からのずれについてより良い洞察を与えるであろう．

(iii)　SMIRNOV (1944) の定理　$S_{1n_1}(x)$ と $S_{2n_2}(x)$ を，連続な分布関数をもつ母集団からの大きさ n_1, n_2 の2つの独立な標本にに基づく2つの経験分布関数とする．$n = n_1n_2/(n_1+n_2)$ とし

$$D^+_{n_1n_2} = \sup_{-\infty<x<\infty} (S_{1n_1}(x) - S_{2n_2}(x))$$

$$D_{n_1n_2} = \sup_{-\infty<x<\infty} |S_{1n_1}(x) - S_{2n_2}(x)|$$

とおく．このとき〔$n_1/n_2 \to c$ (定数) として，$n_1, n_2 \to \infty$ とすると〕

$$\lim_{n_1,n_2\to\infty} P(\sqrt{n} D^+_{n_1n_2} < \lambda) = \begin{cases} 1-\exp(-2\lambda^2) & \lambda > 0 \\ 0 & \lambda \leqslant 0 \end{cases}$$

$$\lim_{n_1,n_2\to\infty} P(\sqrt{n} D_{n_1n_2} < \lambda) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2\lambda^2) & \lambda > 0 \\ 0 & \lambda \leqslant 0 \end{cases}$$

を得る．

この定理の証明については Gihman (1952) を見よ．統計量 $\sqrt{n} D_{n_1n_2}$ は2つの標本が同一母集団からのものであるという仮説の検定に用いることができる．このパーセント点は，極限分布が一致するから，(ii) の $\sqrt{n} D_n$ のそれと同じである．これの代替としては，**6c. 2** で論じた類似標本の一様性を検定するための χ^2 検定がある．これら2つの検定の明白な差異は，やはりわれわれの検出したい，ずれの型の違いにある．

6f. 2　標本分位点の漸近分布

分布関数 F の第 p 分位点は

$$P(x \leqslant \xi_p) \geqslant p, \qquad P(x \geqslant \xi_p) \geqslant q \tag{6f. 2. 1}$$

なるような任意の点 ξ_p として定義される．ここに $q = 1-p$ で，この条件は

$$F(\xi_p) \leqslant p, \qquad F(\xi_p) + P(x = \xi_p) \geqslant p \tag{6f. 2. 2}$$

を意味する．分布関数 $F(x)$ が狭義の単調増加関数でなければ，(6f. 2. 1) や (6f. 2. 2) をみたす ξ_p の値が多数あることがある．

n 個の観測値の第 p 分位点も同じように定義される．それは，($\hat{\xi}_p$ 以下の観測値の数) ≥ $[np]$ で，($\hat{\xi}_p$ 以上の観測値の数) ≥ $[nq]$ であるような値 $\hat{\xi}_p$ である．このとき (6f. 1. 1) で定義した量によって，順序統計量を用いると，観測値の関数 $\hat{\xi}_p$ を次のように定義できることが容易にわかる．

$$\begin{aligned} \hat{\xi}_p(x_1, \cdots, x_n) &= x_{([np]+1)} && np\text{ が整数でないとき} \\ &= \text{閉区間 } [x_{(np)}, x_{(np+1)}] \text{ 内の任意の値} && np\text{ が整数のとき} \end{aligned} \tag{6f. 2. 3}$$

ここに [・] は・を超えない最大の整数を表す．

以下に $p \neq 1$ のときの $\hat{\xi}_p$ の漸近分布に関する2つの命題を考察する．

(i)　(6f. 2. 1) が成り立つような ξ_p の値がただ1つ存在するならば，確率1で $n \to \infty$ のとき $\hat{\xi}_p \to \xi_p$ である．

定義より任意の $\varepsilon > 0$ にたいして，(6f. 2. 2) を用いると

$$F(\xi_p + \varepsilon) \geqslant F(\xi_p) + P(x = \xi_p) \geqslant p \tag{6f. 2. 4}$$

である．いま $F(\xi_p + \varepsilon) = p$ とすると，仮定に反して，$\xi_p + \varepsilon$ も条件 (6f. 2. 1) をみたす．したがって $F(\xi_p + \varepsilon) > p$ となり，グリヴェンコの定理により，確率1で $\lim S_n(\xi_p$

$+\varepsilon) > p$ となる．一方 $S_n(\xi_p + \varepsilon) > p$ は $\hat{\xi}_p \leqslant \xi_p + \varepsilon$ を意味する．したがって，十分大きなnについて確率1で $\hat{\xi}_p \leqslant \xi_p + \varepsilon$ となる．不等式 $\hat{\xi}_p \geqslant \xi_p - \varepsilon$ についても同様の結果が成り立つ．εは任意であるから，求める結果が得られたことになる．

（ii） $F(x)$ は x について連続な確率密度関数 $f(x)$ をもつものとする．さらに，ξ_p は一意で，$f(\xi_p) > 0$ とする．このとき

$$\sqrt{n}(\hat{\xi}_p - \xi_p) \xrightarrow{L} X \sim N\left(0, \frac{p(1-p)}{[f(\xi_p)]^2}\right) \qquad (6f.2.5)$$

が成り立つ．

$\hat{\xi}_p < x$ となる確率は x より小さい観測値の個数が np 以上になる確率と同じで，それは2項確率の和，

$$\sum_{m}^{n} \frac{n!}{r!(n-r)!}[F(x)]^r[1-F(x)]^{n-r}$$

$$= \frac{n!}{(m-1)!(n-m)!}\int_0^{F(x)} t^{m-1}(1-t)^{n-m}dt \qquad (6f.2.6)$$

に等しい．ここで，np が整数なら $m = np$，np が整数でなければ $m = [np] + 1$ とする．2項確率の部分和と不完全ベータ関数とが等しいことを示す (6f.2.6) は容易に証明できる．$\hat{\xi}_p$ の確率密度関数は (6f.2.6) をxについて微分すれば得られ，次式となる．

$$\frac{n!}{(m-1)!(n-m)!}[F(x)]^{m-1}[1-F(x)]^{n-m}f(x) \qquad (6f.2.7)$$

$y = F(x)$ とすれば $dy = f(x)dx$ である．変数 y の確率密度は

$$\frac{n!}{(m-1)!(n-m)!}y^{m-1}(1-y)^{n-m} \qquad (6f.2.8)$$

である．ここで新しい変数

$$z = \frac{\sqrt{n}(y-p)}{\sqrt{pq}}, \qquad y = p\left(1 + z\sqrt{\frac{q}{np}}\right)$$

を導入する．zの確率密度は2つの項の積

$$\frac{1}{\sqrt{n}}\frac{n!}{(m-1)!(n-m)!}p^{m-\frac{1}{2}}q^{n-m+\frac{1}{2}} \times (1 + z\sqrt{q}/\sqrt{np})^{m-1}(1 - z\sqrt{p}/\sqrt{nq})^{n-m} \qquad (6f.2.9)$$

として得られる．(6f.2.9) の対数をとり階乗にたいしてはスターリングの近似を用いると，zを含まない第1項の対数は，$n \to \infty$ のとき

$$\log\left(\frac{1}{\sqrt{n}}\frac{n!}{(m-1)!(n-m)!}p^{m-\frac{1}{2}}q^{n-m+\frac{1}{2}}\right) \to \text{定数}$$

となる．また，zを含む項の対数は，$n \to \infty$ で

$$(m-1)\log\left(1 + \frac{z\sqrt{q}}{\sqrt{pn}}\right) + (n-m)\log\left(1 - \frac{z\sqrt{p}}{\sqrt{qn}}\right) \to -\frac{z^2}{2}$$

となる．したがって，(6f.2.9) の $n \to \infty$ のときの極限は（定数）$\times e^{-z^2/2}$ となり，これ

は正規分布の確率密度関数である．よって，シェッフェの定理 [(xv), **2c.4**] により z の漸近分布は $N(0, 1)$ となる．よって $\sqrt{n}(y-p)$ の漸近分布は $N(0, pq)$ となる．

ところで $y = F(\hat{\xi}_p)$ つまり $\hat{\xi}_p = F^{-1}(y)$ で，$d\hat{\xi}_p/dy = [f(\hat{\xi}_p)]^{-1}$ である．$y = p$ での導関数の値は $[f(\xi_p)]^{-1}$ であるから，[(i), **6a.2**] を用いると，$\sqrt{n}(\hat{\xi}_p - \xi_p)$ の漸近分布は $N(0, pq/[f(\xi_p)]^2)$ となる．

特に，中央値の漸近分布は正規分布で，その平均は母中央値μ，漸近分散は

$$\frac{[f(\mu)]^{-2}}{4n} \qquad (6f.2.10)$$

である．したがって，もとの分布が正規分布ならば，中央値のところでの正規分布の縦座標は $1/\sigma\sqrt{2\pi}$ であるから，正規分布の平均の推定値としての標本中央値の漸近分散は $\pi\sigma^2/2n$ となる．n個の観測値の平均 $\bar{x}$ の分散は σ^2/n であるから，平均の漸近分散と中央値の漸近分散の比は $2/\pi$ であることがわかる．（よって推定量の効率に関する前の定義によれば，母平均の推定量としての中央値の効率は $2/\pi$，すなわち6 %程度ということになる．）

$\hat{\xi}_p$ の正確な分布関数である (6f.2.6) を導く際に

$$P(\hat{\xi}_p < x) = P(nS_n(x) \geqslant np) \qquad (6f.2.11)$$

という等式を用いた．ただし，$S_n(x)$ は (6f.1.2) で定義した2項変量である．Weiss (1970) は，正確な確率分布 (6f.2.7) を経ずに，この (6f.2.11) の関係を用いて，直接に $\hat{\xi}_p$ の漸近分布を求めた．(6f.2.11) で $x = \xi_p + t/\sqrt{n}$ とおき，$n \to \infty$ とすれば

$$P(\sqrt{n}(\hat{\xi}_p - \xi_p) < t)$$

$$= P\left\{\frac{\sqrt{n}}{\sqrt{pq}}[S_n(\xi_p + t/\sqrt{n}) - F(\xi_p + t/\sqrt{n})] > \frac{\sqrt{n}}{\sqrt{pq}}[p - F(\xi_p + t/\sqrt{n})]\right\}$$

$$\stackrel{a}{=} P\left\{\frac{\sqrt{n}}{\sqrt{pq}}[S_n(\xi_p) - F(\xi_p)] \geqslant -\frac{tf(\xi_p)}{\sqrt{pq}}\right\}$$

$$\to P\left\{X \geqslant -\frac{tf(\xi_p)}{\sqrt{pq}}\right\} = P\left\{X < \frac{tf(\xi_p)}{\sqrt{pq}}\right\} \qquad (6f.2.12)$$

となる．ここで2項変数 $S_n(\xi_p)$ に中心極限定理 [(i), **2c.5**] を用いれば $X \sim N(0, 1)$ となり，これがまさに (6f.2.5) の結果である．

(6f.2.12) を導くさいに

$$\sqrt{n}\{[S_n(\xi_p + t/\sqrt{n}) - S_n(\xi_p)] - [F(\xi_p + t/\sqrt{n}) - F(\xi_p)]\} \xrightarrow{P} 0$$

という結果を用いたことに注意せよ．これは左辺の期待値は零で，その（2項）分散は $n \to \infty$ で $\to 0$ となることからわかる．さらに $p = F(\xi_p)$ に注目すると，F についての仮定のために $n \to \infty$ のとき

$$\sqrt{n}[p - F(\xi_p + t/\sqrt{n})] \to -tf(\xi_p)$$

となることも用いている．

$\hat{\xi}_p$ の漸近的性質についての興味ある論文は Bahadur (1966) と Kiefer (1967) による．

6g　統計量の変換

6g.1　一　般　公　式

[(i), 6a. 2]で，$\{T_n\}$，$n=1, 2, \cdots$ が統計量の列とすると，

$$\sqrt{n}(T_n-\theta) \xrightarrow{L} X \sim N[0, \sigma^2(\theta)]$$

$$\Rightarrow \sqrt{n}[g(T_n)-g(\theta)] \xrightarrow{L} X \sim N(0, [g'(\theta)\sigma(\theta)]^2)$$

であることを示した．ただし，g は1階の導関数をもつ関数で，$g'(\theta) \neq 0$ とする．

ここで関数 g として

$$g'(\theta)\sigma(\theta)=c \qquad (\theta \text{ とは無関係に}) \tag{6g. 1. 1}$$

になるものを選べば，変換した統計量 $g(T_n)$ の漸近分散は θ とは無関係になるであろう．微分方程式 (6g. 1. 1) を関数 g について解くと，次を得る．

$$\frac{dg}{d\theta}=\frac{c}{\sigma(\theta)} \quad \text{つまり，} \quad g=c\int\frac{d\theta}{\sigma(\theta)}. \tag{6g. 1. 2}$$

この (6g. 1. 2) の結果は，6g. 2 から 6g. 4 までの節の多くの重要な統計量に適用される．これらの節では変換した統計量の使い方を詳細に述べる．

6g. 2　ポアソン変量の平方根変換

x がポアソン変量で

$$E(x)=\mu, \qquad V(x)=\mu$$

とすれば，$\mu \to \infty$ で $(x-\mu)/\sqrt{\mu} \xrightarrow{L} X \sim N(0, 1)$ である．このとき $g(x)-g(\mu)$ の漸近分布が $N(0, c^2)$ で，c は μ とは無関係な定数になるような関数 g を求めたい．(6g. 1. 2) の公式を使うと，c を適当に選んで

$$g(\mu)=\int\frac{c\,d\mu}{\sqrt{\mu}}=\sqrt{\mu} \tag{6g. 2. 1}$$

とできる．変数を変換して $\sqrt{x}$ とすると，μ が大きいときのこの漸近平均と漸近分散は

$$\sqrt{\mu} \quad \text{と} \quad \frac{1}{4}$$

となる．Anscombe (1948) は b を適当な定数として変換 $\sqrt{x+b}$ をつくると，この変換が理論的にすぐれていることを示した．いま $(x-\mu)=t$，$(\mu+b)=\mu'$ とおき，テイラー展開を用いて

$$\sqrt{x+b}=\sqrt{\mu'}\left\{1+a_1\frac{t}{\mu'}-a_2\left(\frac{t}{\mu'}\right)^2+\cdots+(-1)^s a_{s-1}\left(\frac{t}{\mu'}\right)^{s-1}+\cdots\right\} \tag{6g. 2. 2}$$

を得る．ここに，

$$a_s=(-1)^{s+1}\frac{(-1)(-3)\cdots(-2s+3)}{s!2^s}$$

である．ポアソン分布の積率は

$$E(t)=0, \quad E(t^2)=\mu, \quad E(t^3)=\mu, \quad E(t^4)=3\mu^2+\mu, \quad \cdots$$

であるから，展開式 (6g. 2. 2) の両辺の期待値をとって

$$E(\sqrt{x+b})=\sqrt{\mu+b}-\frac{1}{8\sqrt{\mu}}+\frac{24b-7}{128\mu\sqrt{\mu}}+\cdots$$

$$V(\sqrt{x+b})=\frac{1}{4}\left\{1+\frac{3-8b}{8\mu}+\frac{32b^2-52\ \ +17}{32\mu^2}+\cdots\right\}$$

を得る．ここで $b=\frac{3}{8}$ にとると

$$E\left(\sqrt{x+\frac{3}{8}}\right)=\sqrt{\mu+\frac{3}{8}}-\frac{1}{8\sqrt{\mu}}+\frac{1}{64\mu\sqrt{\mu}}+\cdots$$

$$V\left(\sqrt{x+\frac{3}{8}}\right)=\frac{1}{4}\left(1+\frac{1}{16\mu^2}+\cdots\right)$$

となる．$\sqrt{x+\frac{3}{8}}$ の分散は，その展開式の第2項が $0(1/\mu^2)$ であるから，$\sqrt{x}$ の分散よりも安定である．

6g. 3　2項割合の平方根の逆正弦変換

2項割合 r/n の期待値は π で分散は $\pi(1-\pi)/n$ である．式を解いて変換

$$g(\pi)=\int\frac{c}{\sqrt{\pi(1-\pi)}}d\pi=\sin^{-1}\sqrt{\pi}, \qquad c \text{ を適当に選んで} \tag{6g. 3. 1}$$

$$\text{漸近分散}\left(\sin^{-1}\sqrt{\frac{r}{n}}\right)=\frac{1}{4n}$$

を得る．Anscombe (1948) は，分散がより安定する，これより少しすぐれた変換は

$$\sin^{-1}\sqrt{\left(r+\frac{3}{8}\right)\div\left(n+\frac{3}{4}\right)} \tag{6g. 3. 2}$$

であり，その漸近分布は $1/(4n+2)$ であることを示した．n が大きいときには，簡単なほうの変換 $\sin^{-1}\sqrt{r/n}$ を用いることができ，中ぐらいの大きさの標本にたいしては変換 $\sin^{-1}\sqrt{(r+\frac{3}{8})/(n+\frac{3}{4})}$ を用いればよい．

例　R. A. Fisher (1949) は家ねずみの2つの遺伝子座（猫背と野ねずみ色）の間の次の組換価を発見した．このデータは戻し交雑〔雑種第一代をその一方の親と交配すること〕から得られたものであるので，組換価の推定値は全子孫にたいして組換えの起こったものの割合である．

表 6g. 3α　野ねずみ色遺伝子座が異質接合子をもつ 20 組の両親について観測された組換価

	A^YA^L	A^YA	A^Ya^t	A^Ya	A^LA	
♀	$\frac{12}{194}$	$\frac{5}{118}$	$\frac{12}{235}$	$\frac{10}{146}$	$\frac{2}{78}$	
♂	$\frac{9}{128}$	$\frac{9}{126}$	$\frac{6}{160}$	$\frac{16}{243}$	$\frac{7}{214}$	
	A^La^t	A^La	Aa^t	Aa	a^ta	計
♀	$\frac{9}{182}$	$\frac{4}{210}$	$\frac{10}{231}$	$\frac{8}{178}$	$\frac{11}{159}$	$\frac{83}{1731}$
♂	$\frac{7}{213}$	$\frac{4}{144}$	$\frac{13}{218}$	$\frac{3}{159}$	$\frac{13}{238}$	$\frac{87}{1843}$

表 6g. 3α で分母の数字は組換価を計算するのに寄与した動物の数を示す．

性の差異　各異質接合子について組換価に性の差異があるか？　異質接合子 A^YA^L を考えると，次の 2 × 2 分割表ができる．

	組換え	古い結合	計
♀	12	182	194
♂	9	119	128

独立性検定の χ^2 の値は 0.0905 で自由度は 1 である．10 種の異質接合子についての χ^2 の和は 4.827 で自由度 10 である．自由度 10 の χ^2 が 4.827 を越える確率は大きく，性による差異がないことを示している．

異質接合子間の差異　性の差異は無視できるのだから，性についてはデータを併合し異質接合子の型を 1 つの属性，結合の性質（古いのと新しいのと）を別の属性とする2 × 10 分割表をつくることができる．

一様性の検定の χ^2 は自由度 9 で 16.315 であり，これは 5％水準で，各種の異質接合子たいして組換価に差異のあることを示している．

性と異質接合子との交互作用　この例は性の差異についてこれ以上解析を進めるには不適当である．そこで10種の交配の一部または全部で性による差が存在したとしよう．このときには，性の差異がそのすべての場合について同じであるかどうか，という次の問題が提起される．すなわち，性と異種接合子の種類の間に交互作用があるかどうかを検定する必要が生じる．これには，角変換を行った上で分散分析を適用すればよい．観測した各比率 p について，$p=\sin^2\phi$ となる角度 ϕ を求める．R. M. M. または F. Y. の表のように ϕ が度数で与えられるときは，ϕ の分散は $8100/n\pi^2$，近似的には，$820.7/n$ である．20個の

角度と必要な計算ステップを表 6g. 3β に示す．

表 6g. 3β　交互作用の計算

	A^YA^L	A^YA	A^Ya^t	A^Ya	A^LA	
♀	14.4	11.8	13.0	15.1	9.2	
♂	15.3	15.4	11.2	14.9	10.4	
d (差)	−0.9	−3.6	1.8	0.2	−1.2	
$w=n_1n_2/(n_1+n_2)$	77.12*	60.93	95.19	91.20	57.16	
dw	−69.41	−219.35	171.34	18.24	−68.59	
$d^2w=d(dw)$	62.47	789.65	308.41	3.65	82.31	
	A^La^t	A^La	Aa^t	Aa	a^ta	計
♀	12.7	7.9	11.9	12.2	15.2	
♂	10.4	9.6	14.2	7.9	13.5	
d	2.3	−1.7	−2.3	4.3	1.7	
w	98.14	85.42	112.16	83.98	95.32	856.62
dw	225.72	−145.21	−257.97	361.11	162.04	177.92
d^2w	519.16	246.86	593.33	1552.79	275.47	4434.10

10 種の異質接合子の各々における性の差異を検定するための χ^2 は自由度 10 で

$$\sum d^2w \div 820.7 = 4434.10 \div 820.7 = 5.40$$

であり，これは，さきに直接 χ^2 分析で求めた値 4.827 よりすこし大きい．自由度 10 の χ^2 から，全体としての性の差に起因する自由度 1 の χ^2

$$\frac{(\sum dw)^2}{\sum w} \div 820.7 = \frac{(177.92)^2}{856.62} \div 820.7 = 0.05$$

を引く．その残り

$$5.40 - 0.05 = 5.35$$

は自由度 9 の χ^2 となり，性と異質接合子の型との間の交互作用の検定に用いられる．この交互作用の χ^2 も性の差異による χ^2 も有意でない．このような場合には，種々の型の異質接合子の間に見られる差異は，性別のデータを合計して調べられる．しかしながら，以下では，性の差はあるが交互作用はないと仮定する場合の適切な検定法を例示しよう．

性の差を除去した後の異質接合子の差異　自由度 $20-1=19$ の全体の χ^2 は

$$\left\{\sum n\phi^2 - \frac{(\sum n\phi)^2}{\sum n}\right\} \div 820.7$$

で，この和は 20 個の角度についてとる．それ以降の計算は表 6g. 3γ に示すが，そこでの和はすべての異質接合子についてとる．♀と♂ごとに求めた χ^2 の和は

表 6g.3γ 異質接合子間の差

	$\sum n$	$\sum n\phi$	$\sum n\phi^2$	$\sum n\phi^2-(\sum n\phi)^2/\sum n = 820.7\chi^2$	χ^2
♀	1731	21, 470. 9	274, 666. 21	8, 346. 43	10. 17
♂	1843	22, 699. 4	290, 576. 08	10, 997. 81	13. 40
全体	3574	44, 170. 3	565, 242. 29	19, 351. 02	23. 58

$$10.17+13.40=23.57$$

でその自由度は 18 である．これから自由度 9 の交互作用成分 5. 35 を差引くと，性の差を除去した後の異質接合子の差を検定するための残りの自由度 9 の χ^2 は

$$23.57-5.35=18.22$$

となって，これは有意である．この値は少し違ったやり方でも求められる．異質接合子の差を無視したときの性の χ^2 は，表 6g.3γ の $\sum n\phi$ と $\sum n$ の列から数値をとって

$$820.7\chi^2=\frac{21,470.9^2}{1731}+\frac{22,699.4^2}{1843}-\frac{44,170.3^2}{3574}$$

$$\chi^2=0.01$$

として求められる．すると，異質接合子に対する正当な χ^2 は，全体から 0.01 と交互作用平方和を引いて求められる． これより $23.58-0.01-5.35=18.22$ となり，さきに得たものと同じになる．

表 6g.3δ χ^2 の分解

	自由度	χ^2
性	1	0.05
異質接合子	9	18.22
交互作用	9	5.35
全体	19	23.58

自由度 19 の全体の χ^2 は有意であり*)，全体として差があることを示している．χ^2 の各成分を加えても全体の χ^2 とはならないのは，各比率の分母が等しくないからである．

6g.4 相関係数の $\tanh^{-1}$ 変換

積率相関係数 r の漸近分散は

$$\frac{(1-\rho^2)^2}{n} \qquad (6g.4.1)$$

である．漸近分散の式に含まれる未知母数 ρ を取り除くような変換を求めるために，公式

*) $\chi^2=23.58$ は 10%水準でも有意ではない．全体としては差がうすめられる．［訳注］

(6g.1.2) を適用して

$$g(\rho)=\int\frac{c}{1-\rho^2}d\rho=\tanh^{-1}\rho \qquad c \text{ を適当に選んで}$$

$$\text{漸近分散 } (\tanh^{-1}r)=\frac{1}{n}$$

を得る．$\tanh^{-1}r$ の分布の積率を調べる．いま

$$\zeta=g(\rho)=\frac{1}{2}\log\frac{(1+\rho)}{(1-\rho)}=\tanh^{-1}\rho$$

$$z=g(r)=\frac{1}{2}\log\frac{(1+r)}{(1-r)}=\tanh^{-1}r$$

と定義する．

$z-\zeta=x$ とおけば，x の分布は r の分布から導かれる．z の 4 次までつの積率はフィッシャーにより求められ，のち Gayen (1951) によって改良された．

$$E(z)=\tanh^{-1}\rho+\frac{\rho}{2(n-1)}\left\{1+\frac{5+\rho^2}{4(n-1)}+\cdots\right\}$$

$$\mu_2=\frac{1}{n-1}\left\{1+\frac{4-\rho^2}{2(n-1)}+\frac{22-6\rho^2-3\rho^4}{6(n-1)^2}+\cdots\right\}$$

$$\mu_3=\frac{\rho^3}{(n-1)^3}+\cdots$$

$$\mu_4=\frac{1}{(n-1)^2}\left\{3+\frac{14-3\rho^2}{n-1}+\frac{184-48\rho^2-21\rho^4}{4(n-1)^2}+\cdots\right\}$$

これら 4 つの積率の展開式を用いて．β_1,β_2 を求めると

$$\beta_1=\frac{\rho^6}{(n-1)^3}+\cdots$$

$$\beta_2=3+\frac{2}{n-1}+\frac{4+2\rho^2-3\rho^4}{(n-1)^2}+\cdots$$

となる．

あまり大きくない n についても，β_1 と β_2-3 は小さいから，$z-\zeta$ は近似的に

$$\text{平均}=\frac{\rho}{2(n-1)}$$

$$\text{分散}=\frac{1}{n-1}+\frac{4-\rho^2}{2(n-1)^2}\fallingdotseq\frac{1}{n-3} \qquad (6g.4.2)$$

の正規分布をすると考えてよいことがわかる．

所与の ρ についての検定 28 個の独立な観測値の対からなる標本において，相関係数は 0.6521 であることがわかった．このような値は，母相関係数 ρ が 0.7211 であるような母集団から得られるか？ 相関係数の変換を適当な数値表 (F. Y. または R. M. M.) を用いて行い

$$z=\frac{1}{2}\log_e\frac{1+r}{1-r}=0.7790$$

$$\text{平均の } z=\frac{1}{2}\log_e\frac{1+\rho}{1-\rho}+\frac{\rho}{2(n-1)}$$

$$= 0.9100 + \frac{0.7211}{54} = 0.9233$$

を得る．正規偏差量は

$$\sqrt{n-3}(z - \text{平均の } z) = \sqrt{28-3}(0.7790 - 0.9233)$$
$$= 5(-0.1443) = -0.7215 \qquad (6g.4.3)$$

となる．絶対値が 0.7215 を越える確率は大体 45% であるから仮説は棄却できない．

平均の z についての補正項 $\rho/2(n-1)$ は，n が大きければ重要ではない．この項を使えば，確率はより正確に得られる．

2つの相関係数の差の検定　観測数が n_1 と n_2 の2つの標本の相関係数を r_1 と r_2 とする．これらの値は，2つの標本が母相関係数の等しい2つの母集団からとられたという仮説と矛盾しないか．いま

$$z_1 = \frac{1}{2}\log_e\frac{1+r_1}{1-r_1}, \qquad z_2 = \frac{1}{2}\log_e\frac{1+r_2}{1-r_2}$$

とおく．統計量 $z_1 - z_2$ は，ρ を共通の母相関係数とすると，平均が

$$\frac{\rho}{2(n_1-1)} - \frac{\rho}{2(n_2-1)} \qquad (6g.4.4)$$

で，漸近分散が

$$\frac{1}{n_1-3} + \frac{1}{n_2-3} \qquad (6g.4.5)$$

の分布をする．よって，2つの標本が小さくなければ，統計量

$$\frac{z_1 - z_2}{\sqrt{1/(n_1-3) + 1/(n_2-3)}} \qquad (6g.4.6)$$

は正規偏差量（標準正規分布に従う）と考えてよい．

いくつかの相関係数の均一性の検定　大きさ $n_1, \cdots, n_k$ の標本から求めた k 個の相関係数を $r_1, \cdots, r_k$ とする．$\tanh^{-1}$ 変換によって，$r_1, \cdots, r_k$ に対応する量 $z_1, \cdots, z_k$ が求められる．もし，平均の z に含まれる偏りが無視できるならば，相関係数の均一性の検定は z の平均値の均一性検定と同等になる．その計算手順は以下のようになる．

表 6g.4α　相関係数の類似推定値の均一性の検定

標本番号 t	標本の大きさ n	相関係数 r	$\tanh^{-1}r$ $= z$	分散の逆数 $n-3$	$(n-3)z$	$(n-3)z^2$
1	n_1	r_1	z_1	n_1-3	$(n_1-3)z_1$	$(n_1-3)z_1^2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	n_k	r_k	z_k	n_k-3	$(n_k-3)z_k$	$(n_k-3)z_k^2$
	計			N	T_1	T_2

ρ が共通のとき $\tanh^{-1}\rho$ の最良推定値は T_1/N である．均一性の検定に用いる統計量は（6a.2 の (6a.2.12) と (v) を参照）

$$H = T_2 - \frac{T_1^2}{N} \qquad (6g.4.7)$$

であり，これは自由度 $k-1$ の χ^2 として扱える．

例として大きさが，10, 14, 16, 20, 25, 28 の6個の標本から求めた相関係数が 0.318, 0.106, 0.253, 0.340, 0.116, 0.112 であるとしよう．これらは均一と考えられるか？

相関係数 r	（標本の大きさ）-3 $n-3$	z	$(n-3)z$	$(n-3)z^2$
0.318	7	0.3294	2.3058	0.7595
0.106	11	0.1064	1.1704	0.1245
0.253	13	0.2586	3.3618	0.8694
0.340	17	0.3541	6.0197	2.1316
0.116	22	0.1164	2.5608	0.2981
0.112	25	0.1125	2.8125	0.3164
計	95		18.2310	4.4995

$$\frac{T_1}{95} = \frac{18.2310}{95} = 0.191905$$

$$T_2 - T_1\frac{T_1}{95} = 4.4995 - 3.4986 = 1.0009$$

自由度5の χ^2 の値 1.0009 は有意ではなく，したがって，これらの相関係数の推定値は均一と考えられる．

検定における偏りの補正と ρ の最良推定値　標本の大きさが，それほど大きくなくまたほぼ等しいということもなければ，(6g.4.7) で用いた H 統計量に，ある大きさの偏り（検定する仮説とは無関係な）がもちこまれる．この偏りは，z の平均値の式で $\rho/2(n-1)$ の項を無視したことによる．H 統計量にもちこまれる偏りは小さくても，H が有意でないときに ρ の最良推定値にもちこまれる偏りは，その推定値の標準誤差に比べて小さくはないのである．この偏りは，少し異なったやり方で補正できる．

z は，平均が

$$\frac{1}{2}\log_e\frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}$$

で，分散が $1/(n-3)$ の正規偏差量と考えることができるから，k 個の標本から得られる ρ のスコアは

$$S = \sum(n_i-3)\left[\frac{1}{1-\rho^2} + \frac{1}{2(n_i-1)}\right]\left[z_i - \frac{1}{2}\log_e\frac{1+\rho}{1-\rho} - \frac{\rho}{2(n_i-1)}\right] \qquad (6g.4.8)$$

であり，その情報量は

$$\mathscr{I} = E(S^2) = \sum (n_i - 3)\left[\frac{1}{1-\rho^2} + \frac{1}{2(n_i - 1)}\right]^2 \qquad (6g.4.9)$$

である．前項で求めた ρ の値を第1近似と考えると，この値に加えるべき補正量 $\delta\rho$ は

$$\delta\rho = \frac{S_0}{\mathscr{I}_0}$$

によって与えられる．ここで S_0 と $\mathscr{I}_0$ は選ばれた近似値を使って計算した (6g.4.8) と (6g.4.9) の値である．この手順は補正量が無視できるほどになるまで繰り返し続けられる．ρ の最良推定量 $\hat{\rho}$ が得られたなら，母相関係数が等しいという仮説の検定のための自由度 $k-1$ の H 統計量は

$$H = \sum (n_i - 3)\left\{z_i - \frac{1}{2}\log_e \frac{1+\hat{\rho}}{1-\hat{\rho}} - \frac{\hat{\rho}}{2(n_i - 1)}\right\}^2$$

となる．

6h　積率の標準誤差と関連統計量

6h.1　原点まわりの積率の分散と共分散

$(x_1, \cdots, x_n)$ を任意の母集団（確率変数 x についての）からの独立な観測値とする．確率変数 x の，原点まわりおよび平均値まわりの r 次の積率を

$$\nu_r = E(x^r), \qquad \mu_r = E(x - \nu_1)^r$$

で表す，（観測値についての）原点まわり積率と〔平均値による〕補正済の積率（これを中心積率という）は

$$0_r = \sum x_i^r / n, \qquad m_r = \sum (x_i - 0_1)^r \div n$$

である．このとき，容易に

$$E(0_r) = \sum E(x_i)/n = \nu_r$$

$$E(0_r^2) = E\frac{\sum x_i{}^r + 2\sum\sum x_i^r x_j}{n^2}$$

$$= \frac{n\nu_{2r} + n(n-1)\nu_r^2}{n^2}$$

$$V(0_r) = E(0_r^2) - E^2(0_r) = \frac{\nu_{2r} - \nu_r^2}{n}$$

であることがわかる．同様に

$$\mathrm{cov}(0_r, 0_s) = \frac{\nu_{r+s} - \nu_r \nu_s}{n}$$

である．

原点が母平均にとられているときは，これらの式の ν は μ でおきかえられる．

6h.2　平均値まわりの積率の漸近分散と漸近共分散

（観測値の平均値まわりの）r 次の補正済中心積率は

$$m_r = 0_r - \binom{r}{1}0_{r-1}0_1 + \binom{r}{2}0_{r-2}0_1^2 - \cdots (-1)^r 0_1^r$$

である．これは原点に関して不変であるから，0_r を母平均まわりのなまの積率と考えることもできる．ここで m_r の漸近分散を求めるのに δ 法 (6a.2.9) を用いよう．このため期待値 $E(0_1) = 0$，$E(0_i) = \mu_i$ $(i = 2, \cdots, r)$ における m_r の $0_1, \cdots, 0_r$ に関する偏導関数を求める必要がある．それらは

$$\frac{\partial m_r}{\partial 0_r} = 1 \qquad \frac{\partial m_r}{\partial 0_{r-s}} = (-1)^s \binom{r}{s} 0_1^s = 0 \quad (\text{期待値のところで})$$

$$\frac{\partial m_r}{\partial 0_1} = -r0_{r-1} + \binom{r}{2}0_{r-2}(20_1) - \cdots$$

$$= -r\mu_{r-1} \quad (\text{期待値のところで})$$

となる．したがって，次の同等性を得る．

$$m_r - \mu_r \stackrel{a}{=} (0_r - \mu_r) - r\mu_{r-1}(0_1 - \mu_1)$$

$$\text{漸近分散}\ (m_r - \mu_r) = V(0_r) + r^2\mu_{r-1}^2 V(0_1) - 2r\mu_{r-1}\mathrm{cov}(0_r, 0_1)$$

$$= \frac{1}{n}[\mu_{2r} - \mu_r^2 - 2r\mu_{r-1}\mu_{r+1} + r^2\mu_{r-1}^2\mu_2]$$

同様に

$$\text{漸近共分散}\ (m_r m_s) = \frac{1}{n}[\mu_{r+s}\mu_r\mu_s + rs\mu_2\mu_{r-1}\mu_{s-1} - r\mu_{r-1}\mu_{s+1} - s\mu_{r+1}\mu_{s-1}]$$

を得る．特別の場合として次が得られる．

$$\text{漸近分散}\,(m_2) = \frac{1}{n}(\mu_4 - \mu_2^2)$$

$$\text{漸近分散}\,(m_3) = \frac{1}{n}(\mu_6 - \mu_3^2 - 6\mu_2\mu_4 + 9\mu_2^3)$$

$$\text{漸近分散}\,(m_4) = \frac{1}{n}(\mu_8 - \mu_4^2 - 8\mu_5\mu_3 + 16\mu_2\mu_3^2).$$

6h.3　平均値まわりの積率の分散・共分散の正確な式

平均値まわりの積率は，観測値のベキ乗とそれらの積しか含まないから，その分散と共分散の正確な式を求めることができる．しかし，その計算は繁雑でFisher (1928) によって系統的な計算法が明らかにされた．ここでは，統計の応用で必要とする1,2の場合について直接的な計算法を述べる．まず

$$m_2 = 0_2 - 0_1^2, \qquad V(m_2) = E(0_2^2) - 2E(0_2 0_1^2) + E(0_1^4) - [E(m_2)]^2$$

である．一般性を失うことなく，原点が母平均であると仮定して，

$$E(0_2^2) = \frac{\mu_4 + (n-1)\mu_2^2}{n}, \qquad E(0_1^4) = \frac{\mu_4 + 3(n-1)\mu_2^2}{n^3}$$

$$E(0_2 0_1^2) = \frac{\mu_4 + (n-1)\mu_2^2}{n^2}, \qquad E^2(m_2) = \mu_2^2\left(\frac{n-1}{n}\right)^2$$

を計算し，これより

$$V(m_2) = \frac{(n-1)^2}{n^3}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right)$$

$$= \frac{(n-1)^2}{n^3}\left(\kappa_4 + \frac{2n\kappa_2^2}{n-1}\right) \quad \text{（キュミュラントで表して）} \tag{6h. 3. 1}$$

を得る．k_2 を $n(0_2 - 0_1^2)/(n-1)$ で定義すると，(6h. 3. 1) を $(n-1)^2/n^2$ で割ることにより，

$$V(k_2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1} \tag{6h. 3. 2}$$

を得る．同様に

$$\mathrm{Cov}(m_2, 0_1) = \mu_3\frac{(n-1)}{n^2} \tag{6h. 3. 3}$$

$$\mathrm{Cov}(k_2, 0_1) = \frac{\kappa_3}{n} \tag{6h. 3. 4}$$

となる．式(6h. 3. 2) と (6h. 3. 4) を用いて，ポアソン母集団（その母平均と母分散は等しい）からの大きさ n の標本から求めた平均の推定値 $\bar{x}$ と分散の推定値 k_2 との差の有意性を調べる検定をつくることができる．この検定は，標本をとり出した母集団の分散の過大な拡がりや母集団の不均一性を検出するのに有用である．さて

$$V(\bar{x} - k_2) = V(\bar{x}) + V(k_2) - 2\mathrm{Cov}(\bar{x}, k_2)$$

$$= \frac{\kappa_2}{n} + \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1} - \frac{2\kappa_3}{n}$$

$$= \frac{2\mu^2}{n-1} \qquad (\kappa_2 = \kappa_3 = \kappa_4 = \mu \text{ だから})$$

となるが，μ はポアソン分布の母数である．n が大きければ，統計量

$$(\bar{x} - k_2)/\sqrt{V(\bar{x} - k_2)}$$

は，$V(\bar{x} - k_2)$ の式の μ を $\bar{x}$ でおきかえて

$$\sqrt{n-1}\frac{(\bar{x} - k_2)}{\sqrt{2}\,\bar{x}} \tag{6h. 3. 5}$$

と書け，これは標準正規分布の偏差量として扱える．たとえば29個の観測値からなる標本で標本平均と分散 κ_2 はそれぞれ 1.5172 と 1.3300 であったとしよう．このとき (6h. 3. 5) の値は

$$\frac{\sqrt{28}(0.1872)}{\sqrt{2}\,(1.5172)} = 0.462$$

となり，これは小さく，差は有意ではない．

補足と問題

1　4つのポアソン母集団からの大きさ 120, 10，100, 125 の4標本から，平均がそれぞれ 251/120, 323/100, 180/100, 426/125 と得られた．母集団の平均値は等しいか？

［大きさが n_1, n_2, n_3, n_4 の標本から求めた合計を T_1, T_2, T_3, T_4 とする．仮説が正しいなら，$T = T_1 + T_2 + T_3 + T_4$ を与えたときの T_1, T_2, T_3, T_4 の相対分布は，区画の確率が n_1, n_2, n_3, n_4 に比例するような多項分布となり，χ^2 検定を用いることができる．自由度3の χ^2 は

$$\frac{(T_1 - n_1\bar{T})^2}{n_1\bar{T}} + \cdots + \frac{(T_4 - n_4\bar{T})^2}{n_4\bar{T}}$$

で，ここに $\bar{T} = T/(n_1 + n_2 + n_3 + n_4)$ である．この数値は自由度3で 81.12 となることを確かめよ.］

2　環境と遺伝因子の相対的影響を調べる研究で，知能検査得点の間に次のような相関係数の推定値が得られた．必要なら有意性の検定を行ってこれらの数字について論評せよ．

	2人兄弟		双生児	
	別居	同居	別居	同居
相関係数	0.235	0.342	0.451	0.513
標本の大きさ	50	40	45	55

3　ある月のある市における出生児数を推定する調査で，次のようなデータが得られた．調査した450家族のうち，50家族については誕生はなく，100家族について1人生まれ，そのうち25家族は病院で生まれた．病院での出生数が市全体で1000であるとしたとき，その市の総出生児数を推定せよ．その推定値の標準誤差を求めよ．

4　1ルピーの硬貨100枚と半ルピーの硬貨100枚をいっしょに投げたところ130枚の硬貨の表が出た．1ルピーの硬貨には偏りがないと仮定すると，半ルピーの硬貨について，どんなことがいえるか？

5　自由度 n の χ^2 変数について，$\alpha\sqrt{n}\{(\chi^2/n)^\beta + (\gamma/n) - 1\}$ という変換を考える．この統計量の3次までの積率を平均 0，分散1の正規分布の積率と比較して，α, β, γ を求めよ．このとき

$$\alpha = \sqrt{\frac{9}{2}}, \qquad \beta = \frac{1}{3}, \qquad \gamma = \frac{2}{9}$$

に選べば，変換後の統計量は $n \to \infty$ のとき正規分布に急速に近づくことを示せ．

6　次ページの表のデータは，2つの遺伝子 A と B の連鎖があるときに，生れた動物の分布に関するものである．交配の型（異質接合子の性と相）間に，また同じ交配の型内での性の間に，連鎖についての差異があるかどうかを調べよ．

7　次の場合における尤度比検定を求めよ．

(a)　$N(\mu, \sigma^2)$ からの大きさ n の標本に基づき，σ^2 は未知として，仮説 $\mu = 0$ を検定するとき，検定基準はスチューデントの t 検定になることを示せ．

(b)　$N(\mu, \sigma^2)$ からの大きさ n の標本に基づき，μ は未知として，σ^2 がある値であるという仮説を検定するとき．

(c)　$N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ からの大きさ n_1, n_2 の標本に基づき，μ_1, μ_2 は未知として，

異質接合子の性	相	生れた動物の性	表現型 AB	Ab	aB	ab
♂♂	相引	♀	12	13	11	8
		♂	13	15	16	16
	相反	♀	11	13	13	19
		♂	15	10	10	16
♀♀	相引	♀	30	17	20	13
		♂	18	18	20	24
	相反	♀	17	12	13	17
		♂	15	12	11	14

仮説 $\sigma_1^2 = \sigma_2^2$ を検定するとき．

（d） $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$, $\cdots$, $N(\mu_k, \sigma^2)$ からの大きさ $n_1, n_2, \cdots, n_k$ の標本に基づき，共通の σ^2 は未知として，仮説 $\mu_1 = \mu_2 = \cdots = \mu_k$ を検定するとき．この検定は，一元分類 (**4d. 1**) の分散分析の検定になることを示せ．

（e） 最小2乗法理論の一般模型 (**4b. 2**) のもとで線型仮説にたいする尤度比検定を求めよ．この検定は分散分析表の検定と同じになることを示せ．

8　観測値が正規分布に従うという仮定のもとで，スチューデントの t の分布関数は，自由度が無限大になると正規分布に近づくことを示せ．［ヒント：自由度 n の t 分布の確率密度関数を考えよ．ついで $n \to \infty$ のときの極限を求めよ．その極限も密度関数であることを確かめ，シェッフェの定理 [(xv), **2c. 4**] を使え．］

さらに一般的には，$x_1, \cdots, x_n$ を確率変数 x の独立な観測値とし，$E(x) = \mu$, $V(x) = \sigma^2 < \infty$ とする．$t = \sqrt{n}(\bar{x} - \mu)/s$ を定義せよ．$t \xrightarrow{L} X \sim N(0, 1)$ を示せ．［ヒント：中心極限定理により $\sqrt{n}(\bar{x} - \mu)/\sigma \xrightarrow{L} X \sim N(0, 1)$ で $s \xrightarrow{P} \sigma$ である．ここで，**2c. 4** の (x) を使え．］

9　$(T_{1n}, \cdots, T_{kn})$ を k 次元統計量の系列で，$[\sqrt{n}(T_{1n} - \theta_1), \cdots, \sqrt{n}(T_{kn} - \theta_k)]$ の漸近分布は平均零，共分散行列 σ_{ij} の多変量正規分布であるとしよう．n を陽に含む関数 $g(T_{1n}, \cdots, T_{kn}, n)$ を考え，その関数 $g(x_1, \cdots, x_k, n)$ の $x_1, \cdots, x_k$ についての1階偏導関数が存在し，かつ $n \to \infty, x_i \to \theta_i$, $i = 1, \cdots, k$ のとき $\partial g/\partial x_i \to G_i(\theta_1, \cdots, \theta_k)$ とする．このとき

$$v_n^{-1}\sqrt{n}[g(T_{1n}, \cdots, T_{kn}, n) - g(\theta_1, \cdots, \theta_k, n)]$$

の漸近分布は，平均零，分散1の正規分布となる．ただし T_{in}, $i = 1, \cdots, k$ のかわりに母数の真値が代入されるとき，$v_n \neq 0$ としたら，

$$v_n = \sum\sum \frac{\partial g}{\partial T_{in}} \frac{\partial g}{\partial T_{jn}} \sigma_{ij}$$

と書ける．

10　多項分布の観測度数が指定した期待値に一致するという仮説を検定するために，(6e. 2. 4) の基準 S_0 を求めよ．これは **6b. 1** の χ^2 検定と同じになることを示せ．(6e. 2. 3) の W_0 の検定と (6e. 2. 2) の $(-2\log_e \Lambda_0)$ は χ^2 検定とは異なることに注意せよ．同様に適合度検定のための (6e. 3. 6) の基準 S_c は，**6b. 2** の χ^2 適合度検定と同じであるが，W_c と $(-2\log_e \Lambda_c)$ は χ^2 検定とは異なる．

11　いま $X_n \xrightarrow{L} X$ とする．このとき，x のコンパクト集合の上で一様に x の関数として，$g_n(x) \to g(x)$ で，$g(x)$ が連続ならば $g_n(X_n) \xrightarrow{L} g(X)$.

参　考　文　献

Anscombe, F. J. (1948), The transformation of Poisson' binomial and negative-binomial data, *Biometrika* **35**, 246-254.

Bahadur, R. R. (1966), A note on quantiles in large samples, *Ann. Math. Statist.* 37, 577-580.

Chernoff, H. and E. L. Lehmann (1954), The use of the maximum likelihood estimates in χ^2 test for goodness of fit, *Ann. Math. Statist.* **25**, 579-586.

Cochran, W. G. (1954), Some methods for strengthening the common χ^2 tests, *Biometrics* **10**, 417-451.

Doob, J. L. (1949), Heuristic approach to the Kolmogorov Smirnov theorems, *Ann. Math. Statist.* **20**, 393-403.

Fisher, R. A. (1925), *Statistical Methods for Research Workers* (first edition, twelfth edition, 1954), Oliver and Boyd, Edinburgh.

Fisher, R. A. (1928), Moments and product moments of sampling distributions, *Proc. London Math. Soc.* **30**, 199-238.

Fisher, R. A. (1949), A preliminary linkage test with agouti and undulated mice, *Heredity* **3**, 229-241.

Gayen. A. K. (1951), The frequency distribution of the product moment correlation coefficient in random samples of any size drawn from non-normal universes, *Biometrika* 38, 219-247.

Gihman, I. I. (1952), On the empirical distribution function in the case of grouping of observations (in Russian), *Doklady Akademii Nauk. USSR* 82, 837-840.

Glivenko, V. I. (1933), Sulla determinazione empirica delle leggi di probabilita, *Giorn. Ist. Ital. Attauri* **4**, 92.

Kiefer, J. (1967), On Bahadur representation of sample quantiles, *Ann. Math. Statist.* **38**, 1323-1342.

Neyman, J. and E. S. Pearson (1928), On the use and interpretation of certain test criteria for purposes of statistical inference, *Biometrika* 20A, 175-240 and 263-294.

Pearson, Karl (1900), On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Phil. Mag. Ser.* (5) **50**, 157-172.

Smirnov, N. (1939a), On the estimation of the discrepancy between empirical curves of distribution for two independent samples, *Bull. Math. Univ. Moscow* 2, 3-14.

Smirnov, N. (1939b), Sur les ecarts de la courbe de distribution empirique (in Russian, French summary), *Rec. Math.* N. S. (*Mat Sborn.*) **6**, 3-26.

Wald, A. (1943), Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans. Am. Math. Soc.* **54**, 426-482.

Weiss, Lionel (1970), Asymptotic distribution of quantiles in some nonstandard cases, *Nonparametric Techniques in Statistical Inference*, 343-348.

Yates, F. (1934), Contingency tables involving small numbers and the χ^2 test, *J. Ryo. Statist. Soc.* (*suppl.*) **1**, 217-235.

第7章　統計的推測の理論

はじめに　これまでの章ですでに，仮説検定，区間推定，ベイズ方式，尤度原理，推測確率のような，推測のいくつかの観点にかかわる問題に出会った．これらの他にも，識別（または判別）の問題，決定関数，受入検査，逐次検定と推定，ノンパラメトリック法のような別の観点もある．これらの観点はおのおの，それ自身の論理的基盤をもち，またデータから推測を引きだすのに重要な役割を演じている．統計学に関する最近の文献では，個々の著者の側に，推測のある特定の観点を選び，それを他に優先して主張する傾向がでてきた．しかし，このことは同一の推測方式があらゆる場合に適用可能とか，適切であることはありえないという単純な事実から，明らかに正当とは認められない．

この章は，推測のあらゆる重要な観点の理論的な議論を行うことによって，与えられたデータから引き出される推測の本質的な特徴とそのいろいろな形式を読者に伝えることを目的とする．

統計学の危機？　統計的推測は，特殊から一般への帰納的推理の性質をもっている．したがってその目的に用いられる道具（統計的方法）には，いろいろな論争があるのは当然である．生のデータの解析に，ある提案された統計的方式を用いる研究者は，これらの論争や統計的推測の現代の傾向（改善された手法とはかぎらない）を知っているべきである．これらの論争のいくつかは，本書で述べた統計的方法の議論のなかにも簡潔に紹介されている．より深く勉強するためには以下の文献を挙げる．Barnard (1949), Barnard, Jenkins and Winston (1962), Birnbaum (1962), Fisher (1956), Hogben (1957), Jeffreys(1948), Kyburg (1961), Lindley (1953, 1957), Neyman (1961), Savage (1954, 1963)．また Godambe and Sprott (1971) が編集したシンポジウムの報告を見よ．

7a　統計的仮説検定

7a.1　問題の記述

いままでの章で非常に多くの検定方式を述べ，かつ例示してきたので，統計的仮説検定理論にあてた本節は，死体解剖のように見えるかもしれない．歴史的にもそうである．そのう

え，すべての有意性検定を容認できる解として導き出せるような，統計的仮説検定の首尾一貫した理論は存在しないのである．多くの場合，検定基準は直観的な考察から求められねばならないだろう．それにもかかわらず，仮説検定の問題を明白に理解させるような形式の整った理論は重要である． Neyman and Pearson (1933) によって考えられたそのような理論の1つは重要な発展を遂げた．それは，統計的仮説を検定する際の種々の複雑な問題を人々の眼の前に展開し，この章で論じる判別（識別）の問題，逐次解析，決定関数等の一般理論の構築に導いたからである．

Sを実験の結果をあらわす標本空間とし，xをSの任意の要素とする．さらにH_0をSの集合のボレル集合体$\mathscr{B}$上の確率測度を部分的にあるいは完全に規定する仮説(帰無仮説とよぶ）とする．問題は観測値xに基づいてH_0が真であるかどうかを決定することである．

たとえばH_0を銅貨に偏りがないという仮説とすると，n回の独立な試行から観測された成功の数rに基づいてこの仮説を検定しなければならない．この場合の帰無仮説は，rの確率分布を完全に規定している．

また，ある一定数の個体の身長の度数分布が適当な級区間で観測され，その帰無仮説H_0は身長の理論分布が未知の平均と分散をもつ正規分布であると規定したとする．この後者の場合の帰無仮説は，各級区間における確率の正確な数値を規定しないが，その確率を2つの未知母数の特定の関数として規定する．前者の型の仮説を**単純仮説**，後者を**複合仮説**とよぶ．

帰無仮説H_0を検定するのに，つまりH_0を棄却するかどうかを観測したデータに基づいて決定するのに，どのような方式がとられようと2種類の誤りが伴う．すなわち（a）H_0が正しいときにH_0を棄却する誤りで，**第1種の誤り**という．（b） 実際には対立仮説が正しいときに，H_0を棄却しない誤りで，**第2種の誤り**という．

（非確率化）検定方式では，標本空間を2つの領域wと$S-w$に分割し，観測値xが$x\in w$ならばH_0を棄却し，そうでなければH_0を棄却しないときめる． 領域wを棄却域とよぶ．後節で確率化決定方式として知られる，より一般的な検定方式を考えるが，ここでの議論は非確率化方式にかぎることにしよう．

H_0を第1種の誤りの確率が，［与えられた仮説gの下で集合$A\subset S$の確率測度を$P(A|g)$であらわすと］

$$P(w|H_0)=\alpha \tag{7a.1.1}$$

であるような単純仮説とする．ここでαを有意水準とよぶ．特定の対立仮説$h\in H$（Hは対立仮説のクラス）にたいする第2種の誤りの確率は

$$P(S-w|h)=\beta(h) \tag{7a.1.2}$$

である．H上で定義された関数$\gamma(h)=[1-\beta(h)]$を**検出力関数**という．H_0が複合仮説，つまり単純帰無仮説のクラスであるとき，有意水準を

$$\alpha=\sup_{h\in H_0}P(w|h) \tag{7a.1.3}$$

と定義しよう．

ネイマンとピアソンが提起した問題は，与えられた有意水準にたいし第2種の誤りができるかぎり小さく，つまり検出力関数ができるかぎり高くなるような棄却域を定めることである． この問題及びそれに関連したいくつかの問題を解くために，2，3の数学上の補助定理を考えよう．

実際には，Sの点は確率変数Xの実現値とみなされ，$A\subset S$にたいし$P(A|H)=P(X\in A|H)$となる．このとき，S上で定義される関数TはXの関数$T(X)$と書くことができる．$X=x$のときその値は$T(x)$となる．したがって$\{T(X)>\lambda\}\equiv\{x: T(x)>\lambda\}$であり，$P(\{x:T(x)>\lambda\})=P(T(X)>\lambda)$となる．

7a.2　ネイマン＝ピアソンの基本補助定理と一般化

補助定理1.　$f_0, f_1, \cdots$を空間S上で測度vに関し積分可能な関数，wを

$$\int_w f_i dv=c_i \quad (c_i\text{は与えられた定数}) \quad i=1,2,\cdots \tag{7a.2.1}$$

であるような任意の領域とする．さらに，定数$k_1, k_2, \cdots$が存在して，領域w_0のなかでは$f_0\geqslant k_1f_1+k_2f_2+\cdots$で，外では$f_0\leqslant k_1f_1+k_2f_2+\cdots$であるような$w_0$が条件(7a.1.2)をみたすとしよう．すると次式が成り立つ．

$$\int_{w_0}f_0dv\geqslant\int_w f_0dv. \tag{7a.2.2}$$

領域wとw_0の共通部分をww_0であらわす．(7a.2.1) から共通部分を引くと

$$\int_{w-ww_0}f_idv=\int_{w_0-ww_0}f_idv \qquad i=1,2,\cdots \tag{7a.2.3}$$

を得る．次の差を考えよう．

$$\begin{aligned}\int_{w_0}f_0dv-\int_w f_0dv&=\int_{w_0-ww_0}f_0dv-\int_{w-ww_0}f_0dv\\&\geqslant\int_{w_0-ww_0}\sum k_if_idv-\int_{w-ww_0}\sum k_if_idv\end{aligned} \tag{7a.2.4}$$

(7a.2.3) を用いると式 (7a.2.4) は零になり，したがって (7a.2.2) が得られる．

補助定理2（補助定理1の一般化）.　$f_0, f_1, f_2, \cdots$を補助定理1と同様とし，ϕはS上の点関数で，$0\leqslant\phi\leqslant 1$，かつ

$$\int f_i\phi dv=c_i \quad (c_i\text{は与えられた定数}) \quad i=1,2,\cdots \tag{7a.2.5}$$

とする．さらに，次の (a), (b) をみたすϕ^*が存在するとしよう．

（a） 条件 (7a. 2. 5) がみたされる．

（b） $\phi^* = 0$ $\quad f_0 < k_1 f_1 + k_2 f_2 + \cdots$ のとき

$= 1$ $\quad f_0 > k_1 f_1 + k_2 f_2 + \cdots$ のとき

$=$ 任意 $\quad f_0 = k_1 f_1 + k_2 f_2 + \cdots$ のとき

このとき

$$\int f_0 \phi^* dv \geqslant \int f_0 \phi dv. \tag{7a. 2. 6}$$

S_1, S_2, S_3 をそれぞれ $f_0 < \sum k_i f_i$, $f_0 > \sum k_i f_i$, $f_0 = \sum k_i f_i$ であるような領域とする．S_2 では $\phi^* - \phi \geqslant 0$, S_1 では $\phi^* = 0$ だから $(\phi^* - \phi) \leqslant 0$ である．したがって S_1, S_2 のそれぞれで，また明らかに S_3 で

$$f_0(\phi^* - \phi) \geqslant (\sum k_i f_i)(\phi^* - \phi)$$

となり，その結果 S のすべての点でこれが成り立つ．このようにして条件 (7a. 2. 5) を用いると

$$\int_S f_0(\phi^* - \phi) dv \geqslant \int_S (\sum k_i f_i)(\phi^* - \phi) dv = 0$$

となり，結果 (7a. 2. 6) が得られる．

補助定理 3. $f_1, \cdots, f_k$ を S 上で測度 ν に関して積分可能な関数とし，$\phi_i(x) \geqslant 0$ は，$(\phi_1 + \cdots + \phi_k = 1)$ が成り立つものとする．さらに，$f_{i_1} = \cdots = f_{i_r} > f_{i_{r+1}} \geqslant f_{i_{r+2}} \geqslant \cdots \geqslant f_{i_k}$ のとき，$\phi^*_{i_1} + \cdots + \phi^*_{i_r} = 1$ という条件の下で $\phi^*_{i_{r+1}} = \cdots = \phi^*_{i_k} = 0$, $\phi^*_{i_1}, \cdots, \phi^*_{i_r}$ は非負だが任意であるように選ぶとする．このとき

$$\sum_1^k \int f_i \phi_i^* dv \geqslant \sum_1^k \int f_i \phi_i dv \tag{2a. 2. 7}$$

が成り立つ．

$f_{i_1} = \cdots = f_{i_r} > f_{i_{r+1}} \geqslant \cdots \geqslant f_{i_k}$ ならば

$$\sum_1^k f_i \phi_i^* = \frac{(f_{i_1} + \cdots + f_{i_r})}{r} \geqslant \sum f_i \phi_i \tag{7a. 2. 8}$$

となる．(7a. 2. 8) の両辺を積分すれば，結果 (7a. 2. 7) を得る．補助定理 3 を若干変形したものが補助定理 4 であり，容易に証明できる．

補助定理 4. $S_1, \cdots, S_k$ を空間 S を分割した k 個の互いに排反な領域とする．$S_1^*, \cdots, S_k^*$ を

$$x \in S_i^* \Rightarrow f_i(x) \geqslant f_j(x), \quad j = 1, \cdots, k$$

となるように分割された k 個の互いに排反な領域とする．このとき

$$\int_{S_1^*} f_1 dv + \cdots + \int_{S_k^*} f_k dv \geqslant \int_{S_1} f_1 dv + \cdots + \int_{S_k} f_k dv \tag{7a. 2. 9}$$

が成り立つ．

7a. 3 単純仮説 H にたいする単純仮説 H_0

$P(x|H_0)$ と $P(x|H)$ をそれぞれ H_0, H の下での σ 有限測度 ν に関する x での密度とする．問題は

$$\int_w P(x|H_0) dv = \alpha \quad \text{(指定された値)} \tag{7a. 3. 1}$$

$$\int_w P(x|H) dv \text{ が最大} \tag{7a. 3. 2}$$

となるような棄却域 w を定めることである．

解は補助定理 1 により与えられる．最適棄却域 w は (7a. 3. 1) をみたすような k が存在するならば [$f_0 = P(x|H)$, $f_1 = P(x|H_0)$ と選ぶと]

$$\{x : P(x|H) \geqslant kP(x|H_0)\} \tag{7a. 3. 3}$$

で定義される．

確率変数 X を用いると検定 (7a. 3. 3) が

$$T = \frac{P(X|H)}{P(X|H_0)} \geqslant k \tag{7a. 3. 4}$$

と書けることがわかる．H_0 に関する T の分布を定めよう．分布が連続ならば任意の指定された α にたいし

$$P(T \geqslant k|H_0) = \alpha \tag{7a. 3. 5}$$

となるような k が存在する．この検定 $T \geqslant k$ は単純対立仮説 H に依存する．しかし，この棄却域 $T \geqslant k$ が，対立仮説のクラスのなかでそれを誘導するのに用いた特定の対立仮説に依存しない場合がいくつかある．このとき単純仮説 H_0 にたいするそのようなすべての対立仮説に関して一様最強力 U. M. P. (uniformly most powerful) な検定が得られる．いくつかの例を考えよう．

たいていの例では， 空間 S は n 次元ユークリッド実空間 R^n であり，そのとき確率変数 X はベクトル変数 $\mathbf{X} = (x_1, \cdots, x_n)$ となる．確率変数 $x_1, \cdots, x_n$ は母集団からの観測値あるいは確率変数の互いに独立な観測値とみなすことにする．また，$x_1, \cdots, x_n$ における密度に関して積分する際のダミー変数として同じ記号の $x_1, \cdots, x_n$ を用いることもある．

σ が一定で既知であり，μ が実直線のある範囲にある，正規分布の集まり $N(\mu, \sigma^2)$ を考えよう．$x_1, \cdots, x_n$ は確率変数の n 個の独立な観測値で H_0 はその分布を $N(\mu_0, \sigma^2)$ と規定する帰無仮説とする．ただし μ_0 は平均 μ の規定された値である．この例では空間 S は R^n で X は n 次元ベクトルである．対立仮説での μ の値にたいし，最適検定 (7a. 3. 4) はベクトル $(x_1, \cdots, x_n)$ を $\mathbf{x}$ であらわすと $P(\mathbf{x}|\mu) \geqslant kP(\mathbf{x}|\mu_0)$，すなわち

$$\exp\left[-\sum \frac{(x_i-\mu)^2}{2\sigma^2}\right] \geqslant k\exp\left[-\sum \frac{(x_i-\mu_0)^2}{2\sigma^2}\right] \tag{7a. 3. 6}$$

である．対数をとると（7a. 3. 6）は

$$(\mu-\mu_0)\sum x_i \geqslant c \qquad (\text{定数}) \tag{7a. 3. 7}$$

となる．対立仮説のクラスを $\mu>\mu_0$ で定義すると，棄却域（7a. 3. 7）は $\bar{x}\geqslant c_1$ と同一である．ここで c_1 は $P(\bar{x}\geqslant c_1|\mu_0)=\alpha$ となるように選ぶ．c_1 は μ にはよらないから，U. M. P. 検定は $\mu>\mu_0$ なる対立仮説のクラスにたいして存在する．同様に検定 $\bar{x}\leqslant c_2$ は $\mu<\mu_0$ なる対立仮説のクラスにたいして U. M. P. である．しかし $\mu<\mu_0$, $\mu>\mu_0$ なる対立仮説全体のクラスを考えると U. M. P. 検定は存在しない．

帰無仮説 $\mu=\mu_0$ の下で $\bar{x}\sim N(\mu_0, \sigma^2/n)$ であるから，（任意に与えられた α の値にたいして）$P(\bar{x}\geqslant c_1|\mu_0)=\alpha$ となるような c_1 を求めることができる．このように $\mu>\mu_0$ なる対立仮説の集合にたいし，ある指定された有意水準の最適検定は存在する．$\mu<\mu_0$ にたいしても同様である．

n 回のベルヌイ試行において観測された成功の数 r をもとに成功の確率が π_0 であることを検定する問題を考えてみよう．対立仮説での成功の確率の値を π とすると補助定理1あるいは結果（7a. 3. 4）を適用すると棄却域は

$$\pi^r(1-\pi)^{n-r} \geqslant k\pi_0^r(1-\pi_0)^{n-r} \tag{7a. 3. 8}$$

となる．対数をとると，対立仮説のクラスが $\pi>\pi_0$ のとき棄却域（7a. 3. 8）は $r\geqslant r_0$ であることがわかる．r_0 は

$$P(r\geqslant r_0|\pi_0)=\alpha \tag{7a. 3. 9}$$

となるように定めなければならない．r の分布は離散的だから（7a. 3. 9）が成り立つような r_0 は存在しないかもしれない．しかし，

$$P(r\geqslant r_1|\pi_0)=\alpha_1>\alpha>\alpha_2=P(r\geqslant r_2|\pi_0) \tag{7a. 3. 10}$$

となるような r_1 と r_2（$=r_1+1$）が存在する．もちろん，$r<r_1$ ならば仮説は有意水準 α で棄却されない．$r\geqslant r_2$ ならば H_0 を棄却し，$r=r_1$ ならばある選んだ確率 θ で表が出るような銅貨投げの確率化実験を行い，表が出たら仮説を棄却することに決めよう．$\theta=(\alpha-\alpha_2)/(\alpha_1-\alpha_2)$ のとき結局 H_0 を棄却する確率は，$\alpha_2+(\alpha_1-\alpha_2)\theta=\alpha$ となり，これは指定された有意水準の要求をみたしている．

離散分布の場合に（分布関数の不連続性のために）採用した確率化方式は，どのような問題にも一般化されるだろう．観測値 $x\in S$ が与えられたとき，表が出る確率が $\phi(x)$ の銅貨投げを行い，表が出れば仮説 H_0 を棄却することにきめる．その関数 ϕ を検定関数とよぶ．棄却域としての w は w のなかで $\phi=1$, w の外で $\phi=0$ を選ぶことによって得られる特殊な場合である．H にたいして H_0 を検定する問題は

$$\int_S P(x|H_0)\phi(x)dv=\alpha$$

$$\int_S P(x|H)\phi(x)dv \text{ が最大} \qquad (\text{ただし}\ S\ \text{は全空間})$$

となるように ϕ を定めることに帰着する．補助定理2はこの答を与えてくれる．最適検定関数は次のようになる．

$$\begin{aligned}\phi^*(x)&=1 && P(x|H)>kP(x|H_0)\ \text{のとき}\\ &=0 && P(x|H)<kP(x|H_0)\ \text{のとき}\\ &=\text{任意} && P(x|H)=kP(x|H_0)\ \text{のとき}\end{aligned} \tag{7a. 3. 11}$$

（7a. 3. 11）で k が常に存在し，条件 $E[\phi(x)|H_0]=\alpha$［(0, 1) の任意に与えられた値］をみたすことに注意せよ．

これを証明するために確率変数 $Y=P(X|H)/P(X|H_0)$ と H_0 の下での Y の分布を考えよう．Y の分布関数 F が連続ならば，任意の $0<\alpha<1$ にたいし $1-F(y_\alpha)=\alpha$ となる y_α が存在するので $k=y_\alpha$ とできる．他の場合，F は少なくとも左から連続であり［2a の (iii) を見よ］，与えられた α にたいし

$$1-F(y_\alpha)\geqslant\alpha\geqslant 1-F(y_\alpha+0)$$

となるような y_α が存在する．よって $k=y_\alpha$ とできる．観測された Y の値が y_α ならば確率 θ で H_0 を棄却する．ただし θ は方程式

$$1-F(y_\alpha+0)+\theta[F(y_\alpha+0)-F(y_\alpha)]=\alpha$$

から定まる．この解は2項分布の例に見られるように棄却域（7a. 3. 3）が与えられた大きさ α をもつような k が存在しない場合を含んでいることを除けば（7a. 3. 3）で得られた解と実質的に同一である．解（7a. 3. 11）に対応して3つの領域 $w_1, w_2, S-w_1-w_2$ が得られる．$x\in w_1$ ならば H_0 は棄却され，$x\in w_2$ ならば H_0 は棄却されない．$x\in S-w_1-w_2$ ならば（7a. 3. 10）のときのようにある選んだ確率 θ で表が出るように銅貨を投げて決定する．θ は方程式 $P(w_1)+\theta P(S-w_1-w_2)=\alpha$ から定まることがわかるだろう．

このように2項分布について提案された確率化方式は最適であり，対立仮説のクラス $\pi>\pi_0$ にたいし最大の検出力をもつ．しかしこのようにある事象が観察された後に確率化実験を伴うような方式は，収集したデータの解釈をしたり，それから推測を行おうとする科学者にとって恣意的だと思われるかもしれない．いわゆる有意水準を予め指定された値に正確に等しくすべきであるという要求をみたすためだけにこの恣意性を導入したのであるから，彼の意中での留保は正当であると認めねばならない．$\pi=\pi_0$ なる仮説を検定する際，彼は得られた低いほうの水準の α_2 をそのときの状況から判断して採用し $r\geqslant r_2$ ならばいつでも仮説を棄却することにきめるかもしれない．あるいは α_1 が α よりあまり大

きくないならば，棄却域に r_1 も含むように決めるかもしれない．

確率比検定 P. R. T. (probability ratio test) とよばれる検定 (7a. 3. 11) のいくつかの一般的な性質を考えてみよう．

(i) 確率密度関数の分解定理が成り立つとき，P. R. T. は最小十分統計量の関数である．

尤度比は (2d. 3. 2) の分解を用いると十分統計量の関数であるからこの結果は正しい．

(ii) $\alpha<1$ で H と H_0 の下での分布が異なるならば，検定の検出力が有意水準 α より大きいという意味で P. R. T. は不偏である．

すべての x にたいし $\phi(x)=\alpha$ とする検定を考えよう．このとき有意水準は $E(\phi|H_0)=\alpha$ であり，また検出力は $E(\phi|H)=\alpha$ である．すると明らかに $E(\phi^*|H)\geqslant\alpha$ である．もし $E(\phi^*|H)=\alpha<1$ ならばほとんどいたるところ (v) で $P(X|H)=P(X|H_0)$ となり，仮定に反することになる．よって $E(\phi^*|H)>\alpha$ である．

(iii) 確率密度関数が H_0 の下で $p(\cdot|H_0)$, H の下で $p(\cdot|H)$ である確率変数の互いに独立な観測値 $x_1,\cdots,x_n$ を考えよう．

$$\int p(x|H_0)\log\frac{p(x|H)}{p(x|H_0)}dx=-\delta \tag{7a. 3. 14}$$

とする．ただし ((1e. 6. 6) の不等式により) $\delta>0$ である．さらに

$$\frac{P(\mathbf{x}|H)}{P(\mathbf{x}|H_0)}=\frac{p(x_1|H)\cdots p(x_n|H)}{p(x_1|H_0)\cdots p(x_n|H_0)}\geqslant\lambda_n \tag{7a. 3. 15}$$

を第1種の誤りが $\alpha_n\to\alpha$ $(0<\alpha<1)$ となるような尤度比検定の系列とする．このとき

$$\overline{\lim_{n\to\infty}}\frac{\log\beta_n}{n}\leqslant-\delta \tag{7a. 3. 16}$$

となる．ただし β_n は第2種の誤りである．

定義により

$$\beta_n=\int_{w_n^c}p(x_1|H)\cdots p(x_1|H)dv\leqslant\lambda_n(1-\alpha_n) \tag{7a. 3. 17}$$

ここで $w_n^c=S-w_n$ は棄却域 w_n の補集合である．一方 H_0 の下で大数の法則により

$$n^{-1}\sum_{i=1}^{n}\log\frac{p(x_i|H)}{p(x_i|H_0)}\xrightarrow{\text{a.s.}}-\delta \tag{7a. 3. 18}$$

となり，これから $\alpha>0$ に近づく α_n にたいし

$$\overline{\lim_{n\to\infty}}\,n^{-1}\log\lambda_n\leqslant-\delta \tag{7a. 3. 19}$$

が導かれる．(7a. 3. 17) の両辺の対数をとり，n で割ると

$$n^{-1}\log\beta_n\leqslant n^{-1}\log(1-\alpha_n)+n^{-1}\log\lambda_n$$

が得られる．$\alpha_n\to\alpha<1$ だから $n\to\infty$ のとき $n^{-1}\log(1-\alpha_n)\to0$ となり，(7a. 3. 19) を用いると必要な結果

$$\overline{\lim_{n\to\infty}}\,n^{-1}\log\beta_n\leqslant\overline{\lim_{n\to\infty}}\,n^{-1}\log\lambda_n\leqslant-\delta \tag{7a. 3. 20}$$

が得られる．

さらに α_n が1よりはるかに小さく有界となるような任意の棄却域の系列については

$$\underline{\lim_{n\to\infty}}\,n^{-1}\log\beta_n\geqslant-\delta \tag{7a. 3. 21}$$

となることが示されている (くわしくは Rao (1962d) を見よ).

(7a. 3. 20) と (7a. 3. 21) を組合わせると P. R. T. にたいし

$$\lim_{n\to\infty}n^{-1}\log\beta_n=-\delta \tag{7a. 3. 22}$$

であることがわかる．しかし結果 (7a. 3. 22) から $\delta'<\delta$ にたいし $\beta_n<0(e^{-n\delta'})$ が導かれ，したがって，指数的速度で $\beta_n\to0$ となる．

(iv) (7a. 3. 15) の比が <1 ならば H_0 を採択し，他の場合には H_0 を棄却することにしよう．α_n,β_n をこのような方式についての第1種の誤り，第2種の誤りとする．すると $n\to\infty$ のとき $\alpha_n\to0$, $\beta_n\to0$ となる．

この結果は，標本の大きさが無限に大きくなったとき，仮説 H_0 と H_1 の完全な判別が可能になることを意味している．大数の法則により H_0 が真のとき，

$$\frac{1}{n}\sum\log\frac{p(x_i|H)}{p(x_i|H_0)}\xrightarrow{P}-\delta$$

となる．したがって $n\to\infty$ のとき

$$\alpha_n=P\left[\frac{1}{n}\sum\log\frac{p(x_i|H)}{p(x_i|H_0)}>0\middle|H_0\right]\to0$$

となる．対称性により β_n も0に収束する．

Chernoff (1952) は $P(\mathbf{x}|H)/P(\mathbf{x}|H)\geqslant k$ なる型の検定にたいし，α_n,β_n やさらに k に関して最小になるときの $(\beta_n+\lambda\alpha_n)$ の最小値のより正確な推定値を得た．

7a. 4 局所最強力検定

U. M. P. 検定が存在しないとき，すべての対立仮説にたいして最良であるような単一の棄却域はない．しかし，帰無仮説に (ある意味で) 近い対立仮説にたいしては最良であるような棄却域をみつけることができるかもしれない．しかもそのような棄却域は遠く離れた対立仮説にたいしてもよい性質をもつことが期待される．確率密度関数 $P(x|H)$ が，実直線のある区間で値をとる (単一の) 母数 θ の関数 $P(x,\theta)$ であるような特殊な場合を考えよう．帰無仮説 H_0 は母数の値 $\theta=\theta_0$ を規定する．w が $\int_w P(x,\theta_0)dv=\alpha$ となる任意の棄却域ならば θ の関数としての検定の検出力は $\gamma(\theta)=\int_w P(x,\theta)dv$ である．$\gamma(\theta)$ がテイラー展開できるとすると $[\gamma(\theta_0)=\alpha$ と書いて$]$,

$$\gamma(\theta) = \alpha + (\theta - \theta_0)\gamma'(\theta_0) + \frac{(\theta-\theta_0)^2}{2!}\gamma''(\theta_0) + \cdots. \qquad (7a.4.1)$$

H を片側対立仮説 $\theta > \theta_0$ のクラスとすると，局所最強力片側検定を得るためには，$\gamma'(\theta_0)$ を最大にする必要がある．積分符号の下での微分を仮定すると最大にするものは，$\gamma'(\theta_0) = \int_w P'(x, \theta_0)dv$ である．

（i）w を

$$\gamma(\theta_0) = \int_w P(x, \theta_0)dv = \alpha \qquad (7a.4.2)$$

であるような任意の棄却域 w_0 を $\{x : P'(x, \theta_0) \geqslant kP(x, \theta_0)\}$ なる棄却域で，条件（7a.4.2）が w_0 にたいしてもみたされるような k が存在するとする．このとき

$$\int_{w_0} P'(x, \theta_0)dv \geqslant \int_w P'(x, \theta_0)dv. \qquad (7a.4.3)$$

結果 (7a.4.3) は補助定理1を適用すれば得られる．x は標本全体をあらわし，$P(x, \theta)$ は x における密度をあらわしていることに注意せよ．

同様に対立仮説が $\theta < \theta_0$ ならば，局所最強力な（片側）棄却域は

$$\{x : P'(x, \theta_0) \leqslant kP(x, \theta_0)\} \qquad (7a.4.4)$$

である．

対立仮説が両側ならば，$\gamma'(\theta_0) = 0$ なる局所不偏の制約を課す．この場合，局所最強力検定は $\gamma''(\theta_0)$，つまり（7a.4.1）の展開で $(\theta - \theta_0)^2/2$ の係数を最大にすることにより得られる．このような検定を局所最強力不偏検定とよぶ．問題は以下の条件をみたす棄却域 w を定めることである．

$$\gamma(\theta_0) = \int_w P(x, \theta_0)dv = \alpha \qquad (7a.4.5)$$

$$\gamma'(\theta_0) = \int_w P'(x, \theta_0)dv = 0 \qquad (7a.4.6)$$

$$\gamma''(\theta_0) = \int_w P''(x, \theta_0)dv \quad \text{が最大} \qquad (7a.4.7)$$

最適棄却域はふたたび補助定理1を用いると得られる．

（ii）条件（7a.4.5）と（7a.4.6）をみたす k_1, k_2 が存在するならば，（7a.4.7）が最大となるような最適棄却域 w_0 は次式で定義される．

$$\{x : P''(x, \theta_0) \geqslant k_1 P'(x, \theta_0) + k_2 P(x, \theta_0)\} \qquad (7a.4.8)$$

注　(i) と (ii) は，副次的な条件をみたすような k, k_1, k_2 が存在するときに，解 (7a.4.3)，(7a.4.4)，(7a.4.8) の十分性を示すにすぎない．確率密度関数にいくつかの条件をつけて，Dantzig and Wald (1951) はこの種の解の必要性とまた k_1, k_2 の存在を証明した．(i) と (ii) の結果のいくつかの応用例を考えよう．

$x_1, \cdots, x_n$ を $N(\mu, \sigma^2)$ からの n 個の互いに独立な観測値とし，σ が既知と仮定して，対立仮説が両側であるとき，仮説 $\mu = \mu_0$ を検定することを考えよう．

（n 個の観測値を $\mathbf{x}$ であらわせば）

$$\frac{P'(\mathbf{x}, \mu_0)}{P(\mathbf{x}, \mu_0)} = \frac{n(\bar{x} - \mu_0)}{\sigma^2}, \qquad \frac{P''(\mathbf{x}, \mu_0)}{P(\mathbf{x}, \mu_0)} = \frac{-n}{\sigma^2} + \frac{n^2(\bar{x} - \mu_0)^2}{\sigma^4}$$

は容易に証明できる．したがって（7a.4.8）で定義した最適検定（棄却域 w_0）は

$$\frac{n(\bar{x} - \mu_0)^2}{\sigma^2} \geqslant k_1(\bar{x} - \mu_0) + k_2 \qquad (7a.4.9)$$

である．対称性により $k_1 = 0$ ならば不偏性の条件（7a.4.6）がみたされることが推測できる．よって検定（7a.4.9）は

$$\frac{n(\bar{x} - \mu_0)^2}{\sigma^2} \geqslant k_2 \qquad (7a.4.10)$$

と書ける．$[n(\bar{x} - \mu_0)^2/\sigma^2] \sim \chi^2(1)$ だから定数 k_2 は関係（7a.4.10）の確率が与えられた値 α になるように定めることができる．このように，両側対立仮説にたいする検定（7a.4.10）は片側対立仮説にたいする検定（7a.3.7）とは異なることがわかる．

$x_1, \cdots, x_n$ を確率密度関数が $p(\cdot, \theta)$ である確率変数の n 個の互いに独立な観測値とする．$\theta > \theta_0$ にたいする $H_0 : \theta = \theta_0$ の局所最強力片側検定は（7a.4.3）を適用すると

$$\sum_{r=1}^{n} \frac{p'(x_r, \theta_0)}{p(x_r, \theta_0)} \geqslant k \qquad (7a.4.11)$$

である．確率密度関数 $p(\cdot, \theta)$ は，

$$E\left[\frac{p'(x, \theta_0)}{p(x, \theta_0)}\middle|\theta_0\right] = 0, \qquad V\left[\frac{p'(x, \theta_0)}{p(x, \theta_0)}\middle|\theta_0\right] = i(\theta_0) \qquad (7a.4.12)$$

が存在するという正則条件を満足するとしよう．ただし $i(\theta_0)$ は θ_0 でのフィッシャーの情報量であることに注意せよ．n が大きいならば中心極限定理を適用すると，漸近的に

$$\frac{1}{\sqrt{ni(\theta_0)}} \sum_{1}^{n} \frac{p'(x_r, \theta_0)}{p(x_r, \theta_0)} \sim N(0, 1) \qquad (7a.4.13)$$

となる．これから，d_α を $N(0, 1)$ の上側 α 点とすると（(7a.4.11) の）k の近似値は

$$d_\alpha\sqrt{ni(\theta_0)} \qquad (7a.4.14)$$

となる．このようにして n が大きいとき，局所最強力片側検定の問題の一般解を得た．

統計量（7a.4.13）をもとに次の両側検定をつくることができる．

$$\left|n^{-1}\sum \frac{p'(x_r, \theta_0)}{p(x_r, \theta_0)}\right| > \frac{d_{\alpha/2}\sqrt{i(\theta_0)}}{\sqrt{n}} \qquad (7a.4.15)$$

これは局所最強力不偏ではないが，大標本ではいくつかの最適な性質をもつことが示されている (Rao (1947e), Rao and Poti (1946i), Wald (1941))．

7a.5　複合仮説の検定

前に述べたように，帰無仮説 H_0 はそれが標本空間 S 上の確率測度を完全に規定せず，

可能な確率測度のより広範な集合の部分集合の要素として真の確率測度を規定するだけであるとき，複合であるという．正規分布の母数 μ と σ を含む複合仮説のいくつかの例を次にあげる．

(a)　$H_0: \mu \leqslant 0,\ \sigma^2 = 1$

(b)　$H_0: \mu = 0,\ \sigma^2$ は任意

(c)　$H_0: \sigma^2 = \sigma_0^2$ (与えられた値)，μ は任意

(d)　$H_0: \mu \leqslant 0,\ \sigma^2$ は任意.

複合仮説では棄却域 w の大きさは

$$\sup_{h \in H_0} \alpha(h) = \sup_{h \in H_0} P(w|h) \tag{7a.5.1}$$

で定義される．すべての要素 $h \in H_0$ について $\alpha(h) = \alpha$ ならば，棄却域 w は大きさ α をもち，標本空間に相似であるという．任意の与えられた問題で，最適棄却域の選択を相似な棄却域にのみ限定し，相似な棄却域の全体のクラスを定めることは興味がある．次の結果は相似検定の存在と構成に関するものである．

(i)　大きさ α の相似な棄却域 w が存在し，T は H_0 の下で許容しうる測度の族にたいする有界型完備十分統計量であるとする．すると $T = t$ を与えたときの w の条件つき大きさは α である．

棄却域 w の定義関数を次のように定義し，単純仮説 $h \in H_0$ を選ぶ．

$$\chi(x) = 1, \quad x \in w \text{ のとき}$$
$$= 0, \quad x \notin w \text{ のとき}$$

このとき T の十分性から，$h \in H_0$ によらず

$$E[\chi|T = t, h] = \eta(t) \tag{7a.5.2}$$

が導かれる．w は相似な棄却域だから，すべての $h \in H_0$ にたいして

$$\alpha = E[\chi|h] = E_T(E[\chi|T, h]) = E[\eta(T)|h] \tag{7a.5.3}$$

が得られる．T は有界型完備だからほとんどすべての t にたいし $\eta(t) = \alpha$，つまり，$T = t$ を与えたときの w の条件つき大きさは α である．[$f(\cdot)$ が有界で，すべての $h \in H_0$ にたいし $E[f(T)] = 0$ なら，ほとんどすべての t にたいし $f(t) = 0$ であるとき，T は有界型完備であるという.]

(ii)　$h \in H_0$ とし $g \in H$ (対立仮説のクラス) とする．

$$\int_w P(x|h)dv = \alpha \qquad \text{すべての } h \in H_0 \text{ にたいし}$$

$$\int_w p(x|g)dv \text{ が最大,} \qquad \text{ある与えられた } g \in H \text{ にたいし}$$

となるような w を定める問題は，H_0 によって規定される確率測度の族にたいしては十分で有界型完備であり，H のそれらにたいしては必ずしもそうではないような統計量 T が存在するとき，

$$E[\chi|T = t, h] = \alpha \tag{7a.5.4}$$

$$E[\chi|T = t, g] \text{ は最大} \tag{7a.5.5}$$

を満足する定義関数 χ を定める問題と同一となる．

証明は簡単であり (i) で得られた結果に基づく．このように最適で相似な棄却域を定める問題は，仮説 h と g に関して T を与えたときの S 上の条件つき確率測度を考えて，与えられた大きさの最適棄却域を定める問題に帰着する．

(iii)　$P(\cdot|h)$ と $P(\cdot|g)$ を $h \in H_0$ と $g \in H$ に対応した確率密度関数とし，T を H_0 で規定された測度の族にたいする有界型完備十分統計量とする．棄却域 w_0 を考え，そのなかでは

$$P(x|g) \geqslant \lambda(T)P(x|h) \tag{7a.5.6}$$

で，外では

$$P(x|g) \leqslant \lambda(T)P(x|h) \tag{7a.5.7}$$

であるとする．$T = t$ を与えたときの w_0 の条件つき大きさが α であるような $\lambda(T)$ が存在するならば，w_0 は H_0 を検定するための最良で相似な棄却域となる．さらにこの検定は不偏である．

証明は議論が (i) で導かれた性質を満足する相似な棄却域にかぎられているということを除けば **7a.3** で考えた単純対立仮説にたいする単純帰無仮説の問題の証明と同様である．(iii) で T の十分性のみがわかっていて，その有界完備性は必ずしもわかっていないとしても，(7a.5.6) や (7a.5.7) のような w_0 の構成は (i) の性質をみたす相似検定を与える．しかしそのようにして得られた w_0 が最良であるとはいえないであろう．なぜならば T が有界型完備でないとき (i) の性質が成り立たないような相似な棄却域が存在するかもしれないし，そのような棄却域との比較はなされていないからである．(iii) で述べた方法を説明するためにいくつかの例を考えてみよう．

$\mu = \mu_0$ で σ^2 は任意という仮説 H_0 を $N(\mu, \sigma^2)$ からの n 個の互いに独立な観測値 $x_1, \dots, x_n$ に基づいて検定したいとする．σ^2 を母数とみなすと H_0 の下での確率密度関数は

$$(2\pi\sigma^2)^{-n/2} e^{-\Sigma(x_i - \mu_0)^2/2\sigma^2}$$

となる．これは $T = \Sigma(x_i - \mu_0)^2$ が σ^2 にたいして十分であることを示している．特定の仮説 $h \in H_0: \mu = \mu_0,\ \sigma^2 = \sigma_0^2$ と $g \in H: \mu = \mu_1,\ \sigma^2 = \sigma_1^2$ を選ぼう．すると (7a.5.6) と (7a.5.7) を用いると，求めるべき棄却域は

$$\sigma_1^{-n} e^{-\Sigma(x_i - \mu_1)^2/2\sigma_1^2} \geqslant \lambda(T)\sigma_0^{-n} e^{-\Sigma(x_i - \mu_0)^2/2\sigma_0^2} \tag{7a.5.8}$$

となる．対数をとり，$T=\sum(x_i-\mu_0)^2$ と書くと（7a.5.8）は $\mu_1>\mu_0$ ならば $(\mu_1-\mu_0)(\bar{x}-\mu_0)\geqslant\lambda_1(T)\Rightarrow(\bar{x}-\mu_0)>\lambda_2(T)$ となる．この検定は

$$\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sqrt{T}}\geqslant\lambda_3(T) \qquad (7a.5.9)$$

と同等である．$\lambda_3(T)$ は T を与えたときの棄却域（7a.5.9）の条件つき大きさが α になるように定めなければならない．$\sqrt{n}(\bar{x}-\mu_0)/\sqrt{T}$ と T が互いに独立な分布に従うこと，したがって，T を与えたときの $\sqrt{n}(\bar{x}-\mu_0)/\sqrt{T}$ の条件つき分布が $\sqrt{n}(\bar{x}-\mu_0)/\sqrt{T}$ の周辺分布と同一であることが容易にわかる．仮説 H_0 の下での $\sqrt{n}(\bar{x}-\mu_0)/\sqrt{T}$ の分布は，第3章の（3b.1.8）で得られている．その分布は未知の σ にはよらないことがわかる．b_α をこの分布の上側 α 点とすると，棄却域

$$\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sqrt{T}}\geqslant b_\alpha \qquad (7a.5.10)$$

は必要な性質をみたしている．このように，（7a.5.9）で $\lambda_3(T)=b_\alpha$ と選べば結果（iii）の条件はみたされ，したがって（7a.5.9）は $\mu>\mu_0$ なる任意の対立仮説にたいする最良な相似検定となる．次の式に注意しよう．

$$\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sqrt{T}}=\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sqrt{n(\bar{x}-\mu_0)^2+\sum(x_i-\bar{x})^2}}=\frac{1}{\sqrt{t^2+n-1}}$$

ただし $t=\sqrt{n}(\bar{x}-\mu_0)/\sqrt{\sum(x_i-\bar{x})^2/(n-1)}$ は第3章と第4章で紹介したスチューデントの t 統計量である．棄却域（7a.5.9）は t であらわすと

$$\frac{t}{\sqrt{t^2+(n-1)}}\geqslant b_\alpha,\ \text{あるいは同等であるが，}\ t\geqslant t_\alpha$$

と書ける．ここで t_α は自由度 $n-1$ の t 分布の上側 α 点である．

同様に $\mu<\mu_0$ のとき，検定は $t\leqslant-t_\alpha$ である．対立仮説の性質が未知のときは局所最強力不偏な棄却域の概念を適用するであろう．このような検定は $|t|\geqslant t_{\alpha/2}$（$t_{\alpha/2}$ は t 分布の上側 $\alpha/2$ 点）で与えられることが示されるだろう．

読者は $\sigma^2=\sigma_0^2$，μ は任意という仮説にたいする相似検定を導くのに同じ方法を適用することができる．$S=\sum(x_i-\bar{x})^2$ ならば，棄却域は

$$\frac{S}{\sigma_0^2}\geqslant a_1\quad\text{あるいは}\quad\frac{S}{\sigma_0^2}\leqslant a_2$$

である．ここで a_1,a_2 は対立仮説が $\sigma^2>\sigma_0^2$ あるいは $\sigma^2<\sigma_0^2$ に応じた定数である．局所最強力不偏な棄却域を得るためには，まずある対立仮説の値 σ^2 の下での $u=S/\sigma_0^2$ の分布を求める．対立仮説の値 σ^2 にたいする u の確率密度関数は

$$P(u,\sigma^2)=c\cdot(\sigma_0/\sigma)^{n-1}e^{-u\sigma_0^2/2\sigma^2}u^{(n-3)/2}$$

である．次の条件を満足する2つの値 (u_1,u_2) を定めなければならない．

$$\int_{u_1}^{u_2}P(u,\sigma_0^2)du=1-\alpha\quad(\alpha\text{は有意水準}) \qquad (7a.5.11)$$

$$\int_{u_1}^{u_2}\frac{dP(u,\sigma_0^2)}{d\sigma_0^2}du=0\quad(\text{不偏性}) \qquad (7a.5.12)$$

簡単な計算により

$$\frac{dP(u,\sigma_0^2)}{d\sigma_0^2}=\frac{(-1)}{\sigma_0^2}\frac{d}{du}[u^{(n-1)/2}e^{-u/2}]$$

であることが容易に証明できる．これより方程式（7a.5.12）は

$$u_1^{(n-1)/2}e^{-u_1/2}=u_2^{(n-1)/2}e^{-u_2/2} \qquad (7a.5.13)$$

に帰着する．方程式（7a.5.11）は

$$\int_{u_1}^{u_2}\frac{1}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)}e^{-u/2}u^{(n-3)/2}du=\alpha \qquad (7a.5.14)$$

となる．u_1,u_2 の値は方程式（7a.5.13）と（7a.5.14）を解けば得られる．n のいろいろな値と $\alpha=0.05$，0.01 にたいする不偏な棄却限界 (u_1,u_2) の値は R.M.M. の表に与えられている．

$x_1,\cdots,x_n$ を $N(\theta,1)$ からの n 個の互いに独立な観測値とし，H_0 は $\theta\leqslant0$ であるとしよう．H_0 にたいする相似検定を構成してみよう．統計量 $\bar{x}$ は θ にたいして十分であり，θ の全範囲，$-\infty<\theta<\infty$ についてこれがいえる．χ を相似な棄却域 w の定義関数とすると

$$E[\chi|\bar{x},\theta]=\alpha,\quad\theta\leqslant0 \qquad (7a.5.15)$$

が成り立つ．$\bar{x}$ は θ にたいし，その全範囲 $-\infty<\theta<\infty$ について十分であるから，結果（7a.5.15）は $\theta>0$ についてもまた成り立つ．このとき任意の $\theta>0$ にたいする w の検出力は

$$\underset{\bar{x}}{E}\{E[\chi|\bar{x},\theta]\}=E(\alpha)=\alpha$$

となり，θ には無関係に第1種の誤りと同一である．検定が相似な棄却域に基づくための条件はこのように役に立たない検定を与える．一方，H_0 に単純仮説 $\theta=0$ を選び，対立仮説 $\theta>0$ にたいする検定をつくろう．すでに示したように検定基準は $\bar{x}\geqslant\lambda$ であり，λ は $P(\bar{x}\geqslant\lambda|\theta=0)=\alpha$ となるように定められる．このような場合，

$$P(\bar{x}\geqslant\lambda|\theta\leqslant0)\leqslant\alpha,\quad P(\bar{x}\geqslant\lambda|\theta>0)>\alpha$$

となり，（7a.5.1）で定義した大きさの条件はみたされ，検出力はつねに α より大きく，かつ実際に θ の増加関数となるという望ましい性質をもつ．このようにして，複合仮説 $\theta\leqslant0$ の妥当な検定が存在し，これは相似検定よりも優れている．この例は上にあげた原則のどれかに従って導いた検定基準は，その性能（検出力曲線の性質等）をよほど注意深く吟味しないかぎり採用してはならないことを示している．

次に母数が1つの分布族での複合仮説を考えよう．そこでは相似な棄却域が存在しないかもしれない．その母数を θ とすると，次の帰無仮説 H_0 と対立仮説 H が得られる．

	H_0	H
(a)	$\theta \leqslant \theta_0$	$\theta > \theta_0$
(b)	$\theta \leqslant \theta_0$ あるいは $\theta \geqslant \theta_1$	$\theta_0 < \theta < \theta_1$
(c)	$\theta_0 \leqslant \theta \leqslant \theta_1$	$\theta < \theta_0$ あるいは $\theta > \theta_1$

$\alpha(\theta) = E(\phi|\theta)$ とする．ϕ は検定関数である．一般に問題は $\theta' \in H$ のとき，条件

$$\alpha(\theta) \leqslant \alpha \qquad (\theta \in H_0 \text{ のとき}) \tag{7a.5.16}$$

の下で $\alpha(\theta')$ が最大となるように ϕ を定めることである．

この問題は一般に妥当な解を持たないかもしれない．命題 (iv) は密度関数 $P(x,\theta)$ の族が，**単調尤度比**をもつ，すなわち任意の $\theta < \theta'$ にたいして X の分布が異なり比 $P(x,\theta')/P(x,\theta)$ が $T(x)$ の非減少関数であるような実数値関数 $T(x)$ が存在するとき，望ましい解を与える．

(iv) θ が実母数で，確率変数 X の確率密度 $P(x,\theta)$ が $T(x)$ について，単調尤度比をもつものとする．$\theta > \theta_0$ にたいして $H_0 : \theta \leqslant \theta_0$ を検定する問題では

$$\phi^*(x) = \begin{cases} 1, & T(x) > c \text{ のとき} \\ \delta, & T(x) = c \text{ のとき} \\ 0, & T(x) < c \text{ のとき} \end{cases}$$

(ただし c, δ は $E(\phi^*|\theta_0) = \alpha$ となるように定められる) で与えられる U. M. P. 検定が存在する．さらに $\alpha^*(\theta) = E(\phi^*|\theta)$ は $\alpha^*(\theta) < 1$ をみたす点 θ の狭義の増加関数である．

$H : \theta_1 > \theta_0$ にたいする $H_0 : \theta = \theta_0$ を考えよう．最強力検定 (7a.3.11) は棄却域 $P(x|\theta_1) >, =, < kP(x|\theta_0)$ に基づいている．この棄却域は $T(x) >, =, < c$ と同等であり，したがって検定 ϕ^* は $\theta > \theta_0$ にたいして $\theta = \theta_0$ を検定する問題で U. M. P. である．ϕ^* はまた $\theta > \theta_0$ にたいして $\theta \leqslant \theta_0$ を検定する問題においても U. M. P. であることを示そう．

もちろん，この検定は任意の θ' を対立仮説 $\theta'' > \theta'$ にたいして有意水準 $\alpha^*(\theta')$ で検定するとき最強力であり，その場合，**7a.3** の命題 (ii) を用いると $\alpha^*(\theta'') > \alpha^*(\theta')$ が成り立つ．すると $\theta \leqslant \theta_0$ について $\alpha^*(\theta) \leqslant \alpha$ となり，ϕ^* は (7a.5.16) をみたす．ϕ が (7a.5.16) をみたす別の検定，つまり $E(\phi|\theta_0) \leqslant \alpha$ としよう．このとき $\theta > \theta_0$ について $E(\phi|\theta) \leqslant E(\phi^*|\theta)$ となる．もしそうでなければ ϕ^* は有意水準 α で $H_0 : \theta = \theta_0$ を $H : \theta > \theta_0$ にたいして検定するとき最強力検定であるから矛盾が生じる．

単調尤度比をもつ分布族の代表的な例は1母数指数型分布族 (**3b.7** を見よ)

$$P(x,\theta) = a(\theta)b(x)\exp[d(\theta)T(x)] \tag{7a.5.17}$$

である．ここで $d(\theta)$ は θ の範囲で狭義単調な関数である．この分布族は $d(\theta)$ が θ の増加関数か減少関数かにより，$T(x)$ または $-T(x)$ についての単調尤度比をもつ．いくつかのよく知られた，離散分布 (2項，ポアソン，負の2項分布等) や連続分布 (正規，ガンマ分布等) はこの分布族に属することがわかる．分布族 (7a.5.17) に関して命題 (v) が成り立つ．

(v) 分布族 (7a.5.17) について，帰無仮説 $H_0 : \theta \leqslant \theta_1$ または $\theta \geqslant \theta_2$ を対立仮説 $H : \theta_1 < \theta < \theta_2$ にたいして検定する問題では，次式で与えられる U. M. P. 検定が存在する．

$$\phi^*(x) = \begin{cases} 1 & c_1 < T(x) < c_2 \text{ のとき} \\ \delta_i & T(x) = c_i,\ i = 1, 2 \text{ のとき} \\ 0 & T(x) < c_1 \text{ または } T(x) > c_2 \text{ のとき} \end{cases} \tag{7a.5.18}$$

ただし c_i, δ_i は

$$E(\phi^*|\theta_1) = E(\phi^*|\theta_2) = \alpha \tag{7a.5.19}$$

となるように定められる．

$H_0 : \theta = \theta_1$ または $\theta = \theta_2$ を考え，与えられた $\theta_1 < \theta < \theta_2$ について $E(\phi|\theta)$ が条件 $E(\phi|\theta_1) = E(\phi|\theta_2) = \alpha$ の下で最大となるように ϕ を定めよう．この場合 **7a.2** の補助定理が適用でき次式が得られる．

$$\phi^*(x) = \begin{cases} 1 & P(x|\theta) > k_1P(x|\theta_1) + k_2P(x|\theta_2) \text{ のとき} \\ \delta & P(x|\theta) = k_1P(x|\theta_1) + k_2P(x|\theta_2) \text{ のとき} \\ 0 & P(x|\theta) < k_1P(x|\theta_1) + k_2P(x|\theta_2) \text{ のとき} \end{cases} \tag{7a.5.20}$$

(7a.5.20) の $P(x|\theta)$ に特別な形 (7a.5.17) を代入すると検定 (7a.5.20) と (7a.5.18) は同等であることがわかる．[この結果は，(7a.5.20) の3つの領域が $y^a \gtreqless b + cy$ ($y = e^{dT}$) に対応し，したがって (7a.5.18) で規定される T の区間に対応していることを示せば証明できる]．これより T にのみ依存する検定 (7a.5.20) は $H_0 : \theta = \theta_1$ または $\theta = \theta_2$ について U. M. P. である．次に，命題 (iv) で行ったのと同様の論法を用いると，この検定は $H_0 : \theta \leqslant \theta_1$ または $\theta \geqslant \theta_2$ について U. M. P. であることがいえる．

不偏検定　$\theta \in H_0$ のとき $E(\phi|\theta) \leqslant \alpha$ で $\theta \in H$ のとき $E(\phi|\theta) \geqslant \alpha$ であるならばその検定は不偏であるという．すでに **7a.3** で U. M. P. は，θ を正規分布の平均とすると $H_0 : \theta = \theta_0$ を $\theta \neq \theta_0$ にたいして検定する問題では存在しないことを示した．このような場合，不偏性の制約を設け，U. M. P. 検定を見つけよう．命題 (vi) は特定の場合の解を与える．

(vi) $P(x|\theta)$ が1母数指数型分布族に属するものとする．このとき有意水準 α で，(a) $H : \theta \neq \theta_0$ にたいして $H_0 : \theta = \theta_0$ を検定する問題や，(b) $H : \theta < \theta_1$ また

は $\theta > \theta_2$ にたいして $H_0 : \theta_1 \leqslant \theta \leqslant \theta_2$ を検定する問題に U. M. P. U. (U. M. P. 不偏) 検定が存在し，それぞれ次の検定関数で定義される．

$$\phi(x) = \begin{cases} 1 & T(x) < c_1 \text{ または } > c_2 \text{ のとき} \\ \delta_i & T(x) = c_i,\ i = 1, 2 \text{ のとき} \\ 0 & c_1 < T(x) < c_2 \text{ のとき} \end{cases}$$

ただし $c_1, c_2, \delta_1, \delta_2$ は問題 (a) については方程式 $E(\phi|\theta_0) = \alpha$，$E(T\phi|\theta_0) = \alpha E(T|\theta_0)$ により，問題 (b) については $E(\phi|\theta_1) = E(\phi|\theta_2) = \alpha$ により定められる．

詳しい証明については，読者は Lehmann (1959) の第4章を参照せよ．

7a.6　フィッシャー＝ベーレンス問題

$n_1, \bar{x}_1, s_1^2$ と $n_2, \bar{x}_2, s_2^2$ をそれぞれ $N(\mu_1, \sigma_1^2)$ と $N(\mu_2, \sigma_2^2)$ からの独立な観測値の数，標本平均，標本分散とする．σ_1^2 と σ_2^2 が未知で異なるかもしれないとき，これらの統計量に基づいた仮説 $\mu_1 - \mu_2 = \xi$ (与えられた値) の適切な検定は何であろうか？

$\sigma_1 = \sigma_2$ のとき，統計量

$$t = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\, \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s} \tag{7a.6.1}$$

ただし $s^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2)$ は自由度 $n_1 + n_2 - 2$ の t 分布に従い，差 $\mu_1 - \mu_2$ に関する任意の仮説にたいする相似検定を与える．検定 (7a.6.1) は $\mu_1 - \mu_2 = \xi$ (与えられた値)，$\sigma_1 = \sigma_2$ のとき，まず未知の母数にたいする完備十分統計量が存在することから，次に [(iii), **7a.5**] の方法を適用することにより **7a.5** の t 検定と同様に導ける．

$\sigma_1 \neq \sigma_2$ で $\mu_1 - \mu_2 = \xi$ (与えられた値) のときは，未知の母数にたいする完備十分統計量が存在するかどうかはわかっていない．したがって [(iii), **7a.5**] の方法は相似検定の決定には適用できない．しかし相似検定は存在するのだろうか？　この疑問にたいする答を最近リニクが与えた．彼は不完備十分統計量に基づく相似検定が存在することを示した．しかしその検定方式は複雑である．

フィッシャーとベーレンス (Fisher (1935, 1956) を見よ) は $\mu_1 - \mu_2$ の推測分布に基づく検定を提唱した．これは **7a.1** で述べた仮説検定の基準を満足していない．たとえば第1種の誤りはその帰無仮説にたいするすべての対立仮説について指定した値よりも小さくない．

いろいろな近似的な検定が提案されている．Banerji (1960) による検定は，第1種の誤りに関する不等式を正確にみたしている．提唱された検定は

$$|\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)| \geqslant \left(\frac{t_1^2 s_1^2}{n_1} + \frac{t_2^2 s_2^2}{n_2}\right)^{\frac{1}{2}}$$

である．ここで t_i は自由度 $n_i - 1$ の t 分布の上側 $\alpha/2$ 点である．この検定の第1種の誤りは $\leqslant \alpha$ である．他の近似的な検定は Chernoff (1949)，Cochran and Cox (1957)，Linnik (1963)，Welch (1974) を見よ．

7a.7　検定の漸近効率

母数 θ の値が θ_0 である単純仮説を $\theta \neq \theta_0$ や $\theta > \theta_0$ または $\theta < \theta_0$ のような対立仮説にたいして検定する問題を考えよう．データが n 個の観測値から成るとし，第1種の誤りを $\alpha_n(\theta_0)$，対立仮説の値 θ にたいする第2種の誤りを $\beta_n(\theta)$ とする．検出力関数は $\gamma_n(\theta) = 1 - \beta_n(\theta)$ である．すでに述べたように $\alpha_n(\theta_0)$ の値が同じである他の検定と比べ，どの θ についても $\gamma_n(\theta)$ が最大値をもつような検定は普通存在しない，にもかかわらず2つの与えられた検定を比較し，そのどちらかよい方を選ぶようなときがあるだろう．しかし第1種の誤り $\alpha_n(\theta_0)$ が同じ値である2つの検定が与えられたとき，その検出力関数 $\gamma_n^{(1)}(\theta)$，$\gamma_n^{(2)}(\theta)$ が θ のとりうる範囲で θ のある値について，ある種の不等式をみたし，別の θ の値にたいしてはその逆の不等式を満足するような場合があるかもしれない．このような知識はきわめて有用であるが，実際には検出力関数の計算が非常に難かしく，これらの検定の選択ができるようにある基準――なるべくなら簡単に計算できる単純な数値尺度に依存しなければならないだろう．このような方法は有限な標本に基づく検定基準ではうまくいかないことは明らかである．したがって検定の識別が明確にできるであろう大標本の場合を考えよう．

妥当な検定の系列をとると，任意に定めた対立仮説の値 $\theta\ (\neq \theta_0)$ にたいして $n \to \infty$ のとき $\gamma_n(\theta) \to 1$ となる．このような検定を一致性をもつ検定という．一致性がない検定は $n \to \infty$ のとき対立仮説を，それが真のとき，確実に検出できないという意味で実際貧弱な性能をもつことになる．一致性をもつ検定は一定の θ にたいし $n \to \infty$ のとき $\gamma_n(\theta) \to 1$ であるから，検出力関数 $\gamma_n(\theta)$ の極限値は検定を識別する基準としては役立たない．しかし任意の n にたいし，それがいかに大きくとも $\gamma_n(\theta)$ が1より小さくなるような θ_0 にかぎりなく近い θ の対立仮説の値があるだろう．そこである規定した方法で $\theta \to \theta_0$，$n \to \infty$ とするときの $\gamma_n(\theta)$ の性質を調べ，漸近効率のいくつかの尺度をつくろう．

漸近効率の尺度　互いに独立に同一の分布に従う観測値に基づく検定の系列においては，次のようないくつかの効率の尺度が提案されている．

(a)　$e_1 = \lim\limits_{n\to\infty} n^{-1}[\gamma_n'(\theta_0)]^2$　(7a.7.1)

ただし $\gamma_n'(\theta_0)$ は $\gamma_n(\theta)$ の $\theta = \theta_0$ での1階の導関数．

(b)　$e_2 = \lim\limits_{n\to\infty} n^{-1}\gamma_n''(\theta_0)$　(7a.7.2)

ただし $\gamma_n''(\theta)$ は $\gamma_n(\theta)$ の $\theta=\theta_0$ での2階の導関数．

（c） $$e_3=-\lim_{\theta\to\theta_0}\lim_{n\to\infty}\frac{2n^{-1}\log\beta_n(\theta)}{(\theta-\theta_0)^2} \tag{7a.7.3}$$

（d） $$e_4=\lim_{n\to\infty}\gamma_n\left(\theta_0+\frac{\delta}{\sqrt{n}}\right)\quad \text{ある選んだ}\ \delta\ \text{にたいして．} \tag{7a.7.4}$$

e_1 から e_4 までの各尺度は，大標本において $\theta=\theta_0$ での検出力曲線の局所的性質をあらわしている． 検定の系列 $\{T_n\}$ にたいする尺度 e_r の値を $e_r(T_n)$ であらわそう．$e_r(T_{1n})>e_r(T_{2n})$ ならば，大標本では $\{T_{1n}\}$ の検出力曲線が θ_0 の近傍で $\{T_{2n}\}$ のそれよりも高いことを意味している．

各観測値の共通の確率密度にゆるい正則条件をつけると，e_1 から e_4 のすべての尺度はどの検定方式にもよらない量により上に有界となり，これらは容易に計算できる．任意の検定の系列 $\{T_n\}$ について $e_r(T_n)$ の上限が得られるとき，$\{T_n\}$ は尺度 e_r によれば漸近的に最良となる．その上限は証明なしで述べるが，その証明がなされている原論文は引用しておく．実際，上限は (7a.7.1)－(7a.7.4) で lim sup をとったものである．

（a′） $e_1^{1/2}\leqslant(i/2\pi)^{1/2}\exp(-a^2/2)$．ただし i は1つの観測値に含まれる θ に関するフィッシャーの情報量であり，a は $N(0,1)$ の上側 α 点である (Rao (1962d))．

（b′） $e_2\leqslant i$

（c′） $e_3\leqslant i$ (Rao (1962d))

（d′） $e_4\leqslant 1-\Phi(a-\delta\sqrt{i})$，ただし Φ は $N(0,1)$ の分布関数 (Rao (1963a))

これらの尺度の計算は一般的な検定基準にたいしては容易でない．しかし，最も重要なのは，漸近的に正規分布に従う統計量に基づく検定である．このような場合のいくつかの尺度の計算を説明しよう．

e_1 の計算 定義から e_1 は検出力曲線の $\theta=\theta_0$ での傾きの推定値を与えていることがわかる．この尺度は片側検定（すなわち，対立仮説が $\theta>\theta_0$ または $\theta<\theta_0$ であるとき）を比較するのに役立つ．不偏検定では，e_1 の値は0になるであろう．その場合検出力曲線の $\theta=\theta_0$ での曲率の尺度となる e_2 が適切である．

互いに独立に同一の分布に従い，それぞれ $\theta=\theta_0$ で $i=E(d\log p/d\theta)^2<\infty$ となるような確率密度 $p(x,\theta)$ をもつ観測値を考えよう．$x_1,\cdots,x_n$ で最初の n 個の観測値をあらわし，$P(X_n,\theta)=p(x_1,\theta)\cdots p(x_n,\theta)$ とする．すると

$$Z_n=n^{-\frac{1}{2}}\frac{P'(X_n,\theta_0)}{P(X_n,\theta)} \tag{7a.7.5}$$

は漸近的に平均0，分散 i の正規分布に従う．T_n を $\sqrt{n}(T_n-\theta_0)$ の漸近分布が $N[0,\sigma^2(\theta_0)]$ となるような θ の一致推定量とする．次にあげる命題は $\sqrt{n}(T_n-\theta_0)$ に基づく検定の系列にたいして e_1 の計算を与えるものである．

（i） 検定基準の系列を $U_n\geqslant a$ とする．ただし，$U_n=\sqrt{n}(T_n-\theta_0)/\sigma(\theta_0)$ で a は $N(0,1)$ の上側確率が α の点である．さらに U_n と Z_n の漸近的な同時分布は相関係数が ρ である2変量正規分布とする．関数 $P(X_n,\theta)$ について積分符号の下での微分が有効ならば $e_1=(i\rho^2/2\pi)\exp(-a^2)\propto\rho^2$ である．

定義により

$$\begin{aligned}n^{-1/2}\gamma_n'(\theta_0)&=\int_{U_n>a}n^{-1/2}P'(X_n,\theta_0)dv_n=\int_{U_n>a}Z_nP(X_n,\theta_0)dv_n\\&\to\int_{U>a}ZP(Z,U)dZdU,\qquad n\to\infty\ \text{のとき}\end{aligned} \tag{7a.7.6}$$

ただし $P(Z,U)$ は $V(Z)=i$，$V(U)=1$，$\mathrm{cov}(Z,U)=\rho\sqrt{i}$ の2変量正規密度関数である．積分 (7a.7.6) は容易に求められ，

$$\left(\frac{i\rho^2}{2\pi}\right)^{1/2}e^{-a^2/2} \tag{7a.7.7}$$

となる．よって (i) の結果が証明できた．

統計量 T_n が (5c.2.5)，(5c.2.6) の意味で有効ならば $\rho=1$ で e_1 は上限を得る．このように有効推定量が尺度 e_1 では漸近的に最良な検定基準を与える．T_{1n} と T_{2n} が Z_n との漸近的な相関が ρ_1,ρ_2 で (i) の条件を満足する2つの統計量の系列ならば，T_{1n} にたいする T_{2n} の相対効率は $(\rho_2/\rho_1)^2$ だから，効率を推定するときのように (7a.7.7) のかわりに ρ^2 自身で検定の効率を定義できる．ρ^2 を尺度 e_1' と書こう．

e_2,e_3 の計算は難かしく，統計量 T_n の系列にさらに多くの仮定を必要とする．次に e_4 を考えよう．これはピットマンの効率とよばれ，大標本の検定を比較するのにしばしば用いられる．

（ii） T_n をはじめの n 個の観測値に基づく検定統計量とし，棄却域を $T_n\geqslant\lambda_n$ とする．以下の (a)～(c) を仮定する．

（a） $\lim_{n\to\infty}P(T_n\geqslant\lambda_n)=\alpha$（一定値）$>0$

（b） すべての実数 y にたいし次式をみたす正の量 r と関数 $\mu(\theta)$，$\sigma(\theta)$ が存在する．

$$\lim_{n\to\infty}P\left\{n^r\frac{T_n-\mu(\theta_0+\delta n^{-r})}{\sigma(\theta_0+\delta n^{-r})}<y\,\middle|\,\theta_0+\delta n^{-r}\right\}=\Phi(y) \tag{7a.7.8}$$

ただし $\Phi(y)$ は $N(0,1)$ の分布関数．

（c） $\mu(\theta)$ は $\theta=\theta_0$ で正の導関数 $\mu'(\theta_0)$ を持ち $\sigma(\theta)$ は θ_0 で連続である．このとき

$$\lim_{n\to\infty}\gamma(\theta_0+\delta n^{-r})=\Phi\left(\frac{\delta\mu'(\theta_0)}{\sigma(\theta_0)}-a\right) \tag{7a.7.9}$$

となる．ただし a は $N(0,1)$ の上側確率が α の点である．

多くの実際の応用では $r=1/2$ であり，これは e_4 の定義で選ばれている．われわれは，

仮定 (b) をみたすように r の値を選ぶことにより，$\gamma(\theta_0+\delta n^{-r})$ の極限として，e_4 を定義することもできたであろう．さて定義より

$$\begin{aligned}\gamma_n(\theta_0+\delta n^{-r}) &= P(T_n \geqslant \lambda_n | \theta_0+\delta n^{-r}) \\ &= P\left(n^r \frac{T_n-\mu(\theta_0+\delta n^{-r})}{\sigma(\theta_0+\delta n^{-r})} \geqslant n^r \frac{\lambda_n-\mu(\theta_0+\delta n^{-r})}{\sigma(\theta_0+\delta n^{-r})} \middle| \theta_0+\delta n^{-r}\right\} \\ &= \Phi\left(-n^r \frac{\lambda_n-\mu(\theta_0+\delta n^{-r})}{\sigma(\theta_0+\delta n^{-r})}\right)+\varepsilon_n(\delta) \qquad (7a.7.10)\end{aligned}$$

となる．ここで (7a. 7. 8) の収束は δ については一様でないが y については一様であることに注意すると，$n\to\infty$ のとき $\varepsilon_n(\delta)\to 0$ となる．$\delta=0$ を代入すると

$$\gamma_n(\theta_0)=\Phi\left(-n^r\frac{\lambda_n-\mu(\theta_0)}{\sigma(\theta_0)}\right)+\varepsilon_n(0) \qquad (7a.7.11)$$

となる．(7a. 7. 11) の両辺の極限をとると

$$\alpha=\lim_{n\to\infty}\Phi\left(-n^r\frac{\lambda_n-\mu(\theta_0)}{\sigma(\theta_0)}\right)$$

となり，これは

$$n^{-r}\frac{\lambda_n-\mu(\theta_0)}{\sigma(\theta_0)}=a+\eta_n \quad \text{で} \quad n\to\infty \quad \text{のとき} \quad \eta_n\to 0$$

を示している．したがって $\lambda_n=n^r(a+\eta_n)\sigma(\theta_0)+\mu(\theta_0)$ である．これより (7a. 7. 10) の Φ の引数は

$$\begin{aligned}-n^r\frac{\lambda_n-\mu(\theta_0+\delta n^{-r})}{\sigma(\theta_0+\delta n^{-r})} &= \frac{\delta\dfrac{\mu(\theta_0+\delta n^{-r})-\mu(\theta_0)}{\delta n^{-r}}-(a+\eta_n)\sigma(\theta_0)}{\sigma(\theta_0+\delta n^{-r})} \\ &\to \frac{\delta\mu'(\theta_0)-a\sigma(\theta_0)}{\sigma(\theta_0)}=\frac{\delta\mu'(\theta_0)}{\sigma(\theta_0)}-a\end{aligned}$$

となり必要とする結果が証明できた．

（iii） (ii) の仮定 (b) が成り立つための十分条件は

$$\lim_{n\to\infty}P\left\{n^r\frac{T_n-\mu(\theta)}{\sigma(\theta)}<y\middle|\theta\right\}=\Phi(y) \qquad (7a.7.12)$$

が $\theta_0\leqslant\theta\leqslant\theta_0+\eta$ なる θ について一様に成立することである．ただし η は任意の正の数である．

(iii) の結果を証明するのは簡単である．実際，(ii) の (b) のかわりに条件 (7a. 7. 12) を証明するのは簡単である．さらに (ii) または (iii) の条件をみたす検定基準のみを考えるならば，効率は $[\mu'(\theta_0)/\sigma(\theta_0)]^{1/r}$ と定義できる．これは与えられた δ, σ にたいする (7a. 7. 9) の増加関数である．尺度 $[\mu'(\theta_0)/\sigma(\theta_0)]^{1/r}$ を e_4' とよぶ．

標本の相対的大きさの概念　尺度 e_1 と2つの検定の系列 $\{T_{1n}\}$ と $\{T_{2n}\}$ を考えよう．$\gamma'_{in}(\theta_0)$ で，与えられた n にたいする i 番目の検定 $(i=1,2)$ の検出力曲線の傾きをあらわす．与えられた n にたいし，

$$\gamma'_{in}(\theta_0)=\gamma'_{2N_n}(\theta_0) \qquad (7a.7.13)$$

つまり傾きが等しく，$n\to\infty$ のとき $N_n\to\infty$ となるような数 N_n が存在すると仮定しよう．効率 $e_1(T_1)$ と $e_1(T_2)$ は定義より

$$e_1(T_1)=\lim_{n\to\infty}[n^{-\frac{1}{2}}\gamma'_{1n}(\theta_0)]^2, \qquad e_2(T_2)=\lim_{N_n\to\infty}[N_n^{-1/2}\gamma'_{2N_n}(\theta_0)]^2$$

である．しかし (7a. 7. 13) の関係から $e_1(T_1)$ は次式に等しい．

$$\begin{aligned}\lim_{n\to\infty}[n^{-1/2}\gamma'_{1n}(\theta_0)]^2 &= \lim_{n\to\infty}[n^{-1/2}\gamma'_{2N_n}(\theta_0)]^2 \\ &= \lim_{n\to\infty}\left[\frac{n^{-1/2}}{N_n^{-1/2}}\cdot N_n^{-1/2}\gamma'_{2N_n}(\theta_0)\right]^2 \\ &= e_1(T_2)\lim_{n\to\infty}\left(\frac{n^{-1}}{N_n^{-1}}\right) \qquad (7a.7.14)\end{aligned}$$

これより $n\to\infty$ のとき $\lim(n/N_n)=e_1(T_2)/e_1(T_1)$，すなわち大標本では漸近効率の比が，2つの検定が同じ検出力の傾きを持つために必要な標本の大きさの比の逆数の極限に等しい．このようにして，尺度 e_1 の実際に役立つ解釈が得られた．

尺度 e_4' を考えよう．命題 (ii) または (iii) の仮定の下では

$$\lim_{n\to\infty}\gamma_n(\theta_0+\delta n^{-r})=\Phi\left(\frac{\delta\mu'(\theta_0)}{\sigma(\theta_0)}-a\right)$$

であることがわかる．

2つの検定統計量の系列 $\{T_{1n}\}$ と $\{T_{2n}\}$ を考え，T_{1n} と T_{2n} に関連した関数を μ_1, σ_1 と μ_2, σ_2 であらわす．前と同様に n と N_n を

$$\gamma_{1n}(\theta_0+\delta n^{-r})=\gamma_{2N_n}(\theta_0+\delta n^{-r}) \qquad (7a.7.15)$$

となるようにし，$n\to\infty$ のとき $N_n\to\infty$ と仮定する．すると

$$\begin{aligned}\Phi\left(\frac{\delta\mu_1'(\theta_0)}{\sigma_1(\theta_0)}-a\right) &= \lim_{n\to\infty}\gamma_{1n}(\theta_0+\delta n^{-r}) \\ &= \lim_{n\to\infty}\gamma_{2N_n}[\theta_0+\delta N_n^{-r}(n^{-r}/N_n^{-r})] \\ &= \lim_{N_n\to\infty}\gamma_{2N_n}(\theta_0+\delta\lambda^{-r}N_n^{-r}) \qquad (7a.7.16) \\ &= \Phi\left(\frac{\lambda^{-r}\delta\mu_2'(\theta_0)}{\sigma_2(\theta_0)}-a\right) \qquad (7a.7.17)\end{aligned}$$

となる．ただし $n\to\infty$ のとき $\lambda=\lim(n/N_n)$ である．(7a. 7. 16) から (7a. 7. 17) への移行には (ii) の証明で用いた論法に従い注意深く証明することが必要である．式 (7a. 7. 17) から

$$\frac{\delta\mu_1'(\theta_0)}{\sigma_1(\theta_0)}=\frac{\delta\mu_2'(\theta_0)}{\lambda^r\sigma_2(\theta_0)}$$

すなわち

$$\lambda = \lim_{n\to\infty}\frac{n}{N_n} = \left\{\frac{\mu_2'(\theta_0)}{\sigma_2(\theta_0)} \div \frac{\mu_1'(\theta_0)}{\sigma_1(\theta_0)}\right\}^{1/r}$$
$$= \frac{e_4'(T_2)}{e_4'(T_1)}$$

がでる．このように，T_2 の T_1 に関する相対効率は $n \to \infty$ のとき $\gamma_n(\theta_n) \to \gamma$（ある指定された値）$<1$ となるような一連の値 $\theta_n \to \theta_0$ で同じ検出力を持つために必要な標本の大きさの比の逆数に等しい．

Bahadur (1960), Chernoff (1952), Noether (1952) 他による検定の漸近効率の興味ある別の尺度がいくつかある．しかしそれは本節で考えたものより，やや特殊である．検定の漸近効率とあるパラメトリック検定やノンパラメトリック検定へのその応用についてさらに詳しく知るには，Chernoff and Savage (1958), Pitman (1949), Hodges and Lehmann (1956), Hoeffching and Rosenblatt (1955) を参照するとよい．

7b　信　頼　区　間

7b.1　一般的な問題

これまでの章で区間により母数を推定する例は数多くあげた．Neyman (1935, 1937), による一般的な方法で問題を定式化し，満足できる解がどの程度得られるかを調べよう．

x を標本点とし，$\boldsymbol{\theta}$ を標本空間上の確率分布を規定する母数（多次元かもしれない）とする．Θ を $\boldsymbol{\theta}$ のとりうる値の集合とする．与えられた x にたいし $I(x)$ で $\boldsymbol{\theta}$ の値の集合をあらわすことにする．信頼係数が $1-\alpha$ の $\boldsymbol{\theta}$ の信頼集合推定量 I，あるいは，簡単に $1-\alpha$ の信頼集合推定量とは，すべての $\boldsymbol{\theta} \in \Theta$ にたいし

$$P[x : \boldsymbol{\theta} \in I(x) | \boldsymbol{\theta}] = 1 - \alpha \tag{7b.1.1}$$

を満足する集合をいう．信頼集合推定量 I が不偏であるとは，

$$P[\boldsymbol{\theta}_1 \in I | \boldsymbol{\theta}_2] \leqslant 1 - \alpha, \qquad \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta \tag{7b.1.2}$$

のときをいう．I と J が同じ信頼係数をもつ $\boldsymbol{\theta}$ の2つの集合推定量であって，

$$P[\boldsymbol{\theta}_1 \in I | \boldsymbol{\theta}_2] \leqslant P[\boldsymbol{\theta}_1 \in J | \boldsymbol{\theta}_2], \qquad \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta \tag{7b.1.3}$$

のとき，I は J よりも短いという．条件 (7b.1.2) と (7b.1.3) は仮説検定理論における検定の不偏性と検出力に対応している．

実際にわれわれにとって一般的な興味があるのは，個々の母数の区間推定量であって，一般的な集合推定量ではない．しかし理論上はより一般的な集合推定の問題を考える方が簡単である．

7b.2　信頼集合を構成する一般的な方法

集合推定の問題は未知の母数に関する単純仮説の検定の問題に密接に関係している．この2つの問題の関係を定めるいくつかの結果を証明しよう．

（i）　$w_{\mathbf{a}}$ を単純仮説 $\boldsymbol{\theta} = \mathbf{a}$（与えられた値）を検定するための大きさ α の棄却域とし，$w_{\mathbf{a}}^c$ を $w_{\mathbf{a}}$ の相補的な領域とする．すると，与えられた x について $\mathbf{a}$ の集合

$$I(x) = \{\mathbf{a} : x \in w_{\mathbf{a}}^c\} \tag{7b.2.1}$$

は，与えられた x について $\boldsymbol{\theta}$ にたいする $1-\alpha$ の信頼集合である．

$I(x)$ の構造から

$$x \in w_{\mathbf{a}}^c \Leftrightarrow \mathbf{a} \in I(x). \tag{7b.2.2}$$

よって

$$P(w_{\mathbf{a}}^c | \boldsymbol{\theta} = \mathbf{a}) = P(x : \mathbf{a} \in I(x) | \boldsymbol{\theta} = \mathbf{a}). \tag{7b.2.3}$$

(7b.2.3) の左辺の式は $1-\alpha$ なる値をもつから，求める結果の証明ができた．

（ii）　$I(x)$ が $\boldsymbol{\theta}$ の $1-\alpha$ の信頼集合とする．このとき x の値の集合

$$w_{\mathbf{a}}^c = \{x : \mathbf{a} \in I(x)\} \tag{7b.2.4}$$

は仮説 $\boldsymbol{\theta} = \mathbf{a}$ にたいする採択域である．

（iii）　$w_{\mathbf{a}}$ が仮説 $\boldsymbol{\theta} = \mathbf{a}$ を検定する問題の棄却域として不偏であるならば，$w_{\mathbf{a}}$ に基づく集合推定量 I は (7b.1.2) の意味で不偏である．逆に I が (7b.1.2) の意味で不偏ならば，I に基づく棄却域 $w_{\mathbf{a}}$ は仮説 $\boldsymbol{\theta} = \mathbf{a}$ の不偏検定を与える．

(7b.2.2) の同値関係で $\boldsymbol{\theta} = \mathbf{b} \neq \mathbf{a}$ にたいする確率をとるとこの結果が得られる．

（iv）　I が別の集合推定量 J より短いとし，$w_{\mathbf{a}}(I)$，$w_{\mathbf{a}}(J)$ をそれに関連した棄却域とする．このとき $w_{\mathbf{a}}(I)$ に基づいた検定は $w_{\mathbf{a}}(J)$ に基づいた検定よりも強力である．

(7b.2.2) の同値関係から

$$\begin{aligned} P(w_{\mathbf{a}}^c(I) | \boldsymbol{\theta}) &= P(\mathbf{a} \in I | \boldsymbol{\theta}), \qquad \boldsymbol{\theta} \neq \mathbf{a} \\ &\leqslant P(\mathbf{a} \in J | \boldsymbol{\theta}) \qquad (I\ \text{の方が短いから}) \\ &= P(w_{\mathbf{a}}^c(J) | \boldsymbol{\theta}) \end{aligned}$$

（v）　(iv) の逆もまた正しい．

(iv) と (v) の結果は U. M. P. 検定が仮説 $\boldsymbol{\theta} = \mathbf{a}$（任意に与えられた値）にたいして存在するならば最短の信頼集合が存在し，その逆も成り立つことを示している．

一般に $\boldsymbol{\theta}$ に関する単純仮説にたいして妥当な検定基準があるならば，満足すべき集合推定量が得られる．たとえば $N(\mu, \sigma^2)$ からの標本 $(x_1, \cdots, x_n)$ に基づいて，σ^2 が未知のときの $\mu = \mathbf{a}$ にたいする妥当な検定は，棄却域が

$$w_{\mathbf{a}} = \left\{x : \frac{\sqrt{n}\,|\bar{x} - a|}{s} \geqslant k\right\}$$

である スチューデントの t 検定である．ここで k は自由度 $n-1$ の t 分布の上側 $\alpha/2$ 点である．このとき，(7b.2.1) より

$$I(x)=\{a : x\in w_a^c\}$$
$$=\left\{a : \frac{\sqrt{n}\,|\bar{x}-a|}{s}\leqslant k\right\}$$
$$=\left\{a : \bar{x}-\frac{ks}{\sqrt{n}}\leqslant a\leqslant \bar{x}+\frac{ks}{\sqrt{n}}\right\}$$

であり，これから μ の $1-\alpha$ の信頼集合（区間）推定量として

$$\left(\bar{x}-\frac{ks}{\sqrt{n}}, \bar{x}+\frac{ks}{\sqrt{n}}\right)$$

が得られる．

また片側検定

$$w_a=\left\{x : \frac{\sqrt{n}\,(\bar{x}-a)}{s}\geqslant k\right\}\qquad (k\text{ は自由度 } n-1 \text{ の } t \text{ 分布の上側 } \alpha \text{ 点})$$

に対応して $1-\alpha$ の片側信頼区間

$$I(x)=\left\{a : \frac{\sqrt{n}\,(\bar{x}-a)}{s}\leqslant k\right\}$$
$$=\left\{a : a\geqslant \bar{x}-\frac{ks}{\sqrt{n}}\right\}\tag{7b.2.5}$$

が得られることがわかる．

上で考えた2つの例では集合推定量は区間であった．次に，反転の方法 (7b.2.1) が1次元の母数にたいしてさえも，必ずしも区間推定量とならないことを示す例を考えよう．

$x\sim N(\mu,1)$ と $y\sim N(\lambda\mu,1)$ は互いに独立とする．仮説 $\lambda=a$ を検定するために次の検定基準を用いる．

$$\frac{(ax-y)^2}{a^2+1}\sim\chi^2(1)\tag{7b.2.6}$$

大きさ α の棄却域は $(ax-y)^2\geqslant(a^2+1)k$ である．ただし k は $\chi^2(1)$ の上側確率が α の点である．λ の集合推定量は $(ax-y)^2\leqslant k(a^2+1)$，つまり

$$a^2(x^2-k)-2axy+y^2-k\leqslant 0\tag{7b.2.7}$$

となるような a の値の集合である．不等式 (7b.2.7) から導かれる集合は，たとえば (x^2-k) が負で a についての2次式の根が実数ならば，区間にはならないだろう．このことは x と y が任意の値をとることができるので起こりうる．しかし推定値が区間でなくなるような (x,y) の値の集合が得られる確率は小さく，その場合には上に提案された解が採択できる．集合推定量が区間推定量にのみ限定されるならば別の方式が必要となる．

7b.3　θの関数の集合推定量

θ を1次元の母数，I を θ の集合推定量，$g(\theta)$ を θ の任意の関数とする．集合 $I(x)$ の θ の値に対応する $g(\theta)$ の値の集合 $g[I(x)]$ を考えよう．$g(x)$ が θ の1対1関数ならば

$$g(\theta)\in g[I(x)]\Leftrightarrow\theta\in I(x)$$

である．この場合 $g[I]$ は $g(\theta)$ の集合推定量であり，信頼係数は I の θ にたいするのと同じである．一方 $g(\theta)$ が1対1関数でないならば

$$\theta\in I(x)\Rightarrow g(\theta)\in g[I(x)]$$

であり，したがって推定量 $g[I]$ の信頼係数は I のそれよりも高い．しかしこれは望ましい解ではない．このような場合，与えられた信頼係数で $g(\theta)$ の集合推定量を得るには，$g(\theta)$ 自身に関する単純仮説の検定にたいする検定基準を求めることが必要であろう．

$\boldsymbol{\theta}$ を多次元の母数，I を $1-\alpha$ の信頼集合推定量としよう．前と同様に，$g[I(x)]$ を集合 $I(x)$ の $\boldsymbol{\theta}$ の値に対応する $g(\boldsymbol{\theta})$ の値の集合とする．一般に

$$\boldsymbol{\theta}\in I(x)\Rightarrow g(\boldsymbol{\theta})\in g[I(x)]$$

だけはいえる．したがって $g[I]$ は信頼係数が $1-\alpha$ より大きい $g(\boldsymbol{\theta})$ の集合推定量である．読者は **4b.2** であげた同時信頼区間に関する例と議論を参照せよ．

7c　逐　次　解　析

7c.1　ワルドの逐次確率比検定

実際にはすべての調査は継続して行われる．実験は，ある決定を下すのに十分な証拠が集積するまで続けられるか，またはある段階で資源の不足のために，あるいはそれ以上の継続がむだであると判断されるために中止される．どの段階においても，それまでに収集された証拠はその先の調査の計画に利用されるのは普通である．ベンガルにおけるジュートの作付面積についての大規模な調査の有効な設計のために予備情報を集めようとして Mahalanobis (1940) によって導入されたパイロット調査の概念は，逐次解析の典型的な応用例である．これとは少し異なる場で，1ロットまたは1バッチの製品の品質がその小標本を検査することによって判定しようとする受入検査において，Dodge and Roming (1929) は2段階の抜取方式を導入した．r_1 を大きさ n_1 の最初の標本中の不良品の数とする．$r_1\leqslant c_1$ ならばそのロットを合格にし，$r_1\geqslant c_2$ ならばロットを不合格にする．$c_1<r_1<c_2$ ならばさらに大きさ n_2 の標本をとり，両方の標本での不良個数の合計 r が c を超えなければロットを合格にし，そうでなければ不合格にする．

逐次解析の一般理論は第2次世界大戦中に受入検査に関連してワルドによってつくられ

た．本節では Wald (1947) による展開に従って，逐次解析のいくつかの基礎概念と方法を考えよう．

逐次確率比検定　H_0 と H_1 を，必ずしも独立に同一の分布に従わない確率変数の系列 $(x_1, x_2, \cdots)$ に関する2つの対立する仮説とする．最初の m 個の観測値に基づく H_0 と H_1 の下での確率密度を $P(\cdot, \cdots, \cdot \mid H_i)$ $i=0,1$ であらわす．最初の m 個の観測値に基づく尤度比は

$$R_m = \frac{P(x_1, \cdots, x_m \mid H_1)}{P(x_1, \cdots, x_m \mid H_0)} \tag{7c. 1. 1}$$

である．2つの対立する仮説 H_0, H_1 間の決定を下す（あるいは H_0 を対立仮説 H_1 にたいして検定する）ためのワルドの逐次確率比検定 S. P. R. T. (sequential probability ratio test) は次のように定義される．

$0<B<1<A<\infty$ となるような2つの定数 A と B を選び固定する．確率変数 $(x_1, x_2, \cdots)$ を次々に観察する．各段階で尤度比を計算する．第 m 段階で　(a)　$R_m \leqslant B$ ならば抜取りをやめ H_0 を採択する．(b)　$R_m \geqslant A$ ならば抜取りをやめ H_1 を採択する．(c)　$B<R_m<A$ ならば抜取りを続けさらに観測を行う．定数 A と B を S. P. R. T. の境界点とよぶ．

一定の大きさの標本に基づく仮説検定理論でのように2種類の誤りが生じる．H_0 が真のとき，H_0 を棄却する確率を α，H_1 が真のとき H_0 を採択する確率を β とすると，この対 (α, β) は逐次検定の強さとよばれる．

7c. 2　S. P. R. T. のいくつかの性質

S. P. R. T. の性質に関するいくつかの結果を証明しよう．

(i)　強さが (α, β) で境界点が (A, B) の S. P. R. T. が確率1で終わるならば

$$A \leqslant \frac{1-\beta}{\alpha}, \qquad B \geqslant \frac{\beta}{1-\alpha} \tag{7c. 2. 1}$$

である．

w_m を $B<R_i<A$, $i=1, \cdots, m-1$, $R_m \geqslant A$ なる領域とする．H_0 が真のとき，H_0 を棄却する確率は

$$\begin{aligned}\alpha &= \sum_{m=1}^{\infty} \int_{w_m} P(x_1, \cdots, x_m \mid H_0) dv^{(m)} \\ &\leqslant \sum_{m=1}^{\infty} \int_{w_m} A^{-1} P(x_1, \cdots, x_m \mid H_1) dv^{(m)} = A^{-1}(1-\beta)\end{aligned}$$

である．ただし $dv^{(m)} = dx_1 \cdots dx_m$．これから (7c. 2. 1) の最初の不等式が成り立つ．同様にして2番目の不等式も証明される．

(ii)
$$A = \frac{1-\beta}{\alpha}, \qquad B = \frac{\beta}{1-\alpha} \tag{7c. 2. 2}$$

と選んだとき，S. P. R. T. が確率1で終わり，強さが (α', β') ならば

$$\alpha' \leqslant \frac{\alpha}{1-\beta}, \quad \beta' \leqslant \frac{\beta}{1-\alpha}, \quad (\alpha'+\beta') \leqslant \alpha+\beta \tag{7c. 2. 3}$$

である．

(7c. 2. 1) を用い A と B に (7c. 2. 2) の式を代入すると

$$\frac{1-\beta}{\alpha} \leqslant \frac{1-\beta'}{\alpha'}, \qquad \frac{\beta}{1-\alpha} \geqslant \frac{\beta'}{1-\alpha'} \tag{7c. 2. 4}$$

を得る．(7c. 2. 4) の最初の不等式から

$$\alpha' \leqslant (1-\beta') \frac{\alpha}{1-\beta} \leqslant \frac{\alpha}{1-\beta}$$

であることがわかる．同様にして $\beta' \leqslant \beta/(1-\alpha)$ となる．さらに (7c. 2. 4) から

$$\frac{1-\alpha'}{\beta'} \geqslant \frac{1-\alpha}{\beta} \Rightarrow \frac{1-\alpha'-\beta'}{\beta'} \geqslant \frac{1-\alpha-\beta}{\beta}$$
$$\frac{1-\beta'}{\alpha'} \geqslant \frac{1-\beta}{\alpha} \Rightarrow \frac{1-\alpha'-\beta'}{\alpha'} \geqslant \frac{1-\alpha-\beta}{\alpha}$$

を得る．これより $(1-\alpha-\beta)>0$ ならば（これは α と β が小さいとき成り立つ）

$$\frac{\alpha'+\beta'}{1-\alpha'-\beta'} \leqslant \frac{\alpha+\beta}{1-\alpha-\beta} \Rightarrow \alpha'+\beta' \leqslant \alpha+\beta$$

となる．

不等式 (7c. 2. 3) は，α と β が小さいとき，α', β' が α, β に近く，かつ不等式 $\alpha' \leqslant \alpha$, $\beta' \leqslant \beta$ のうちの少なくとも1つが成り立つ $(\alpha'+\beta' \leqslant \alpha+\beta$ だから) ことを示している．一般的には，この不等式は両方とも成り立ち，(7c. 2. 2) のような A, B を特に選べば A, B を正しく選んだときの検定よりもより近迫した検定が導かれるであろう．

(iii)　連続した観測値 $(x_1, x_2, \cdots)$ が互いに独立に同一の分布に従うとする．$z(x) = \log[p(x \mid H_1)/p(x \mid H_0)]$ とする．ただし $p(\cdot \mid H_0)$ と $p(\cdot \mid H_1)$ はそれぞれ H_0 と H_1 の下での1つの観測値 x の確率密度である．さらに S. P. R. T. を用いて決定を下すのに必要な観測値の数（確率変数である）を n であらわす．$P(|z(x)|>0 \mid H)>0$ となる任意の仮説 H（必ずしも H_0 または H_1 でない）に関して次の結果が成り立つ．

(a)　$P(n<\infty)=1$ つまり S. P. R. T. は結局終了する．

(b)　$t_0>0$ とすると，$-\infty<t<t_0$ にたいして $E(e^{tn})<\infty$．

結果 (a) は，得られる観測値の分布が何であっても，$P(|z(x)|>0)>0$ でさえあれば，S. P. R. T. は確率1で終わることを示している．結果 (b) は，n のすべてのモーメントが有限であることを示している．

$z_i = \log[p(x_i \mid H_1)/p(x_i \mid H_0)]$ とすると，$(z_1, z_2, \cdots)$ は互いに独立に同一の分布に従

う確率変数の系列となる．逐次方式は，

$$b=\log B<(z_1+\cdots+z_i)=S_i<\log A=a,\qquad i=1,2,\cdots,$$

つまり $S_i\in(b,a)$ であるかぎり続く．m と k を固定し，r を (m/k) を超えない最大の整数として，$S_k, S_{2k}-S_k, S_{3k}-S_{2k},\cdots,S_{rk}-S_{(r-1)k}$ を考えよう．もし必要な標本の大きさ n が m をこえるならば，$i=1,2,\cdots,m$，特に $i=k,2k,\cdots,rk$ にたいして，$S_i\in(b,a)$ である．したがって $|T_i|=|S_{ik}-S_{(i-1)k}|<(|b|+|a|)=c$，$i=1,\cdots,r$ となり，これから

$$P(n>m)\leqslant P(|T_i|<c,\ i=1,\cdots,r)$$
$$=[P(|T_1|>c)]^r\qquad(T_i\text{ は独立であるから})\qquad(7c.2.5)$$
$$\lim_{m\to\infty}P(n>m)=\lim_{r\to\infty}[P(|T_1|<c)]^r\qquad(k\text{ を一定に保って})$$
$$=0\qquad(P(|T_1|<c)<1\text{ ならば})$$

が導かれる．$P(|z|>0)>0$ だから $P(z>h)$ と $P(z<-h)$ が共に正またはどちらか一方が正であるような定数 h が存在する．次に k を c/h より大きい一定の整数にとると，

$$P(|T_1|>c)=P(|z_1+\cdots+z_k|>c)$$
$$\geqslant P\left(z_i>\frac{c}{k},\ i=1,\cdots,k\right)+P\left(z_i<-\frac{c}{k},\ i=1,\cdots,k\right)$$
$$\geqslant[P(z>h)]^k+[P(z<-h)]^k>0$$

となる．したがって $P(|T_1|<c)=\delta<1$ となり，(iii) の命題 (a) が証明された．

(b) を証明するために次式を考えよう．

$$E(e^{tn})=\sum_{m=1}^{\infty}e^{tm}P(n=m)$$
$$\leqslant\sum e^{tm}P(n>m-1)$$
$$\leqslant\sum e^{tm}\delta^r,\quad r=\left[\frac{m-1}{k}\right],\quad \delta=P(|T_1|<c)\qquad((7c.2.5)\text{ を用いて})$$
$$=\sum e^{tm}\delta^{m/k}\delta^{(r-m/k)}$$
$$\leqslant\delta^{-1/k}\sum e^{tm}\delta^{m/k}=\delta^{-1/k}\sum(e^t\delta^{1/k})^m$$

級数 $\sum(e^t\delta^{1/k})^m$ は $e^t\delta^{1/k}<1$ つまり $t<(-\log\delta)/k=t_0$ ならば収束する．ただし，$t_0>0$ でその値は実際に定めることができることに注意せよ．

(iv) 値 $0,1,2,\cdots$ をとる任意の確率変数 n にたいして次式が成り立つ．

$$E(n)=\sum_{m=1}^{\infty}P(n\geqslant m).\qquad(7c.2.6)$$

定義より，

$$E(n)=P(n=1)+2P(n=2)+3P(n=3)+\cdots$$
$$=P(n=1)+P(n=2)+P(n=3)+\cdots$$

$$+P(n=2)+P(n=3)+\cdots$$
$$+\cdots\cdots\cdots\cdots+$$
$$=P(n\geqslant1)+P(n\geqslant2)+P(n\geqslant3)+\cdots$$

となる．

(v) **一般的な補助定理**　互いに独立に同一の分布に従う確率変数 x の観測値の系列 $x_1,x_2,\cdots$ と，与えられた停止規則をもつ逐次決定方式を考えよう．決定に到達するのに必要な観測値の数を n，x の任意の可測関数を $z(x)$，x の確率分布を規定するある仮説を H とする．このとき

$$E(|z(x)||H)<\infty,\ E(n|H)<\infty\Rightarrow E(S_n|H)=E(z|H)E(n|H).$$

ただし $S_n=z(x_1)+\cdots+z(x_n)$ で，決定を下す段階までの z の値の累積和である．

この問題は仮説検定の問題と考える必要はないが，より一般的な性質の問題で，抜取りの継続や決定を下すための停止についての特定の規則をもつものであることに注意せよ．

次のように y_i を定義する．

$$y_i=\begin{cases}1 & \text{第 } i-1 \text{ 段階までに決定が行われないとき}\\ 0 & \text{第 } i-1 \text{ 段階よりも前に決定が行われたとき}\end{cases}$$

すると y_i は明らかに $x_1,\cdots,x_{i-1}$ のみの関数だから x_i にはよらない．したがって $z_i=z(x_i)$ にはよらない．次の和を考えよう．

$$z_1y_1+z_2y_2+\cdots+z_ny_n+z_{n+1}y_{n+1}+\cdots$$

これは容易に S_n であることがわかる．期待値をとると

$$E(S_n)=E(\sum z_iy_i)=\sum E(z_iy_i)$$
$$=\sum E(z_i)E(y_i)=E(z)\sum E(y_i)$$
$$=E(z)\sum P(n\geqslant i)=E(z)E(n)\qquad((7c.2.6)\text{ を用いて})$$

となる．ここで，E と $\sum$ の交換は，仮定により

$$\sum_1^{\infty}E(|z_iy_i|)=E(|z|)\sum P(n\geqslant i)=E(|z|)E(n)<\infty$$

であるから，正当化できる．

7c.3 S.P.R.T. の効率

観測値の系列 $(x_1,x_2,\cdots)$ が互いに独立に同一の分布に従うとし，1つの観測値 x の H_0 と H_1 の下での確率密度を $p(\cdot|H_0)$，$p(\cdot|H_1)$ とする．[(iii)，7c.2] のように

$$z(x)=\log\left[\frac{p(x|H_1)}{p(x|H_0)}\right],\qquad a=\log A,\ b=\log B$$

とする．これらの条件の下で，S.P.R.T. は標本の大きさを一定とした検定よりも，検定の強さ (α,β) が同じならば，前者の平均検査個数 A.S.N. (average sample number)

が後者の一定の検査個数よりも小さいという意味で優れていることを示そう．さらに他のいかなる逐次検定方式に比べても，S. P. R. T. の A. S. N. は最小である．次の結果を証明しよう．

（i）S. P. R. T. は H_0 の下でも H_1 の下でも確率1で終わる．

(1e. 6. 6) で示したように，確率1で $z(x)=0$ でないならば

$$E[z(x)|H_0]<0$$

である．これから

$$P(z<0|H_0)>0 \text{ つまり } P(|z|>0|H_0)>0$$

が出る．したがって［(iii)，**7c. 2**］の結果 (a) を適用すれば，S. P. R. T. は H_0 が真のとき確率1で終わる．H_1 が真のときも証明は同様である．

（ii）$E(|z||H_0)<\infty$, $E(|z||H_1)<\infty$ とする．このとき H_0, H_1 の下での A. S. N. の近似式は

$$E(n|H_0)\doteqdot\frac{b(1-\alpha)+a\alpha}{E(z|H_0)} \qquad (7c.3.1)$$

$$E(n|H_1)\doteqdot\frac{b\beta+a(1-\beta)}{E(z|H_1)} \qquad (7c.3.2)$$

である．ただし (α,β) は S. P. R. T. の強さである．

［(iii)，**7c. 2**］の (b) から $E(n)<\infty$ であり，［(v)，**7c. 2**］から

$$E(n|H_0)E(z|H_0)=E(S_n|H_0) \qquad (7c.3.3)$$

である．$E(S_n|S_n\leqslant b)\doteqdot b$, $E(S_n|S_n\geqslant a)\doteqdot a$ なる近似を使うと，つまり S. P. R. T. が終わるとき S_n の境界を越える分を無視することにより

$$\begin{aligned}E(S_n|H_0)&=P(S_n\leqslant b)E(S_n|S_n\leqslant b)+P(S_n\geqslant a)E(S_n|S_n\geqslant a)\\&\doteqdot(1-\alpha)b+\alpha a\end{aligned} \qquad (7c.3.4)$$

となる．(7c. 3. 3) と (7c. 3. 4) を組み合わせると (7c. 3. 1) を得る．同様にして (7c. 3. 2) も証明できる．

（iii）確率1で終わる任意の逐次検定方式について次式が成り立つ．

$$E(n|H_0)\geqslant\frac{(1-\alpha)\log\dfrac{\beta}{1-\alpha}+\alpha\log\dfrac{1-\beta}{\alpha}}{E(z|H_0)} \qquad (7c.3.5)$$

(7c. 3. 1) で示したように S. P. R. T. について

$$E(n|H_0)\doteqdot\frac{(1-\alpha)b+\alpha a}{E(z|H_0)} \qquad (7c.3.6)$$

となる．ただし a と b は次のような近似値をもつ．

$$a=\log\frac{1-\beta}{\alpha},\qquad b=\log\frac{\beta}{1-\alpha}$$

結果 (7c. 3. 5) が正しいとしよう．すると (7c. 3. 6) は近似的に (7c. 3. 5) の下限となり S. P. R. T. にたいする A. S. N. がほぼ最小であることを示している．

(7c. 3. 5) を証明するために［(v)，**7c. 2**］の結果

$$E(n|H_0)E(z|H_0)\equiv E(S_n|H_0) \qquad (7c.3.7)$$

を用い，任意の逐次検定にたいし $E(S_n|H_0)$ を計算する．

$$\begin{aligned}E(S_n|H_0)&=P(H_0\text{ を採択する})E(S_n|H_0, H_0\text{ を採択する})\\&\quad+P(H_0\text{ を棄却する})E(S_n|H_0, H_0\text{ を棄却する})\\&=(1-\alpha)E(S_n|H_0, H_0\text{ を採択する})\\&\quad+\alpha E(S_n|H_0, H_0\text{ を棄却する})\end{aligned} \qquad (7c.3.8)$$

イェンセンの不等式 (1e. 5. 6) により

$$\begin{aligned}&E(S_n|H_0, H_0\text{ を採択する})\\&\quad\leqslant\log E(e^{S_n}|H_0,\ H_0\text{ を採択する})\\&\quad=\log\frac{1}{1-\alpha}\sum_1^\infty\int_{w_m}e^{S_m}p(x_1|H_0)\cdots p(x_m|H_0)dv^{(m)}\end{aligned} \qquad (7c.3.9)$$

となる．ただし w_m は H_0 を採択する R^m の領域である．

$$e^{S_m}=\frac{p(x_1|H_1)\cdots p(x_m|H_1)}{p(x_1|H_0)\cdots p(x_m|H_0)}$$

であるから，式 (7c. 3. 9) は次式に帰着する．

$$\log\frac{1}{1-\alpha}\sum_1^\infty\int_{w_m}p(x_1|H_1)\cdots p(x_m|H_1)dv^{(m)}=\log\frac{\beta}{1-\alpha}.$$

同様に $E(S_n|H_0, H_0\text{ を棄却する})<\log[(1-\beta)/\alpha]$．したがって (7c. 3. 8) から

$$E(S_n|H_0)\leqslant(1-\alpha)\log\frac{\beta}{1-\alpha}+\alpha\log\frac{1-\beta}{\alpha}$$

であることがわかる．さらに，$E(z|H_0)<0$，だから式 (7c. 3. 7) から不等式 (7c. 3. 5) が得られる．

7c. 4　逐次検定の経済性の例

H_0 の下で $x\sim N(\theta_0,\sigma^2)$，$H_1$ の下で $x\sim N(\theta_1,\sigma^2)$ とする．ここで $\theta_1>\theta_0$ で σ^2 は既知である．最初の m 個の互いに独立な観測値に基づく尤度比の対数は

$$\begin{aligned}\log R_m&=-\frac{1}{2\sigma^2}\left[\sum(x_i-\theta_1)^2-\sum(x_i-\theta_0)^2\right]\\&=\frac{\theta_1-\theta_0}{2\sigma^2}\left[2\sum_1^m x_i-m(\theta_1+\theta_0)\right]\end{aligned}$$

である．第 m 段階での検定方式は次のようになる．

（a）$\displaystyle\sum_1^m x_i-\frac{m(\theta_1+\theta_0)}{2}\leqslant\frac{b\sigma^2}{\theta_1-\theta_0}$ ならば，H_0 を採択する．

(b) $\sum_1^m x_i - \frac{m(\theta_1+\theta_0)}{2} \geqq \frac{a\sigma^2}{\theta_1-\theta_0}$ ならば，H_0 を棄却する．

(c) $\frac{b\sigma^2}{\theta_1-\theta_0} < \sum_1^m x_i - \frac{m(\theta_1+\theta_0)}{2} < \frac{a\sigma^2}{\theta_1-\theta_0}$ ならば，抜取りを続ける．

ただし $a=\log[(1-\beta)/\alpha]$，$b=\log[\beta/(1-\alpha)]$ である．

A. S. N. を定めるのに次の近似式を用いる．

$$E(n|H_0) \doteqdot \frac{b(1-\alpha)+a\alpha}{E(z|H_0)} \tag{7c. 4. 1}$$

次に

$$z = -\frac{(x-\theta_1)^2-(x-\theta_0)^2}{2\sigma^2}, \qquad E(z|H_0) = -\frac{(\theta_1-\theta_0)^2}{2\sigma^2} \tag{7c. 4. 2}$$

である．α, β, a, b の与えられた値にたいして (7c. 4. 1) の値は (7c. 4. 2) の値を $E(z|H_0)$ に代入すれば計算できる．同様にして

$$E(n|H_1) = \frac{b\beta+a(1-\beta)}{(\theta_1-\theta_0)^2/2\sigma^2} \tag{7c. 4. 3}$$

も計算できる．

同じ強さ (α, β) を持つ標本の大きさが一定の検定を考えよう．有意水準 α で標本の大きさが一定の検定は

$$\sqrt{N}(\bar{x}-\theta_0) \geqq d_\alpha\sigma$$

である．ただし N は標本の大きさで，d_α は $N(0,1)$ の上側 α 点である．次に

$$\begin{aligned}\beta &= P(\sqrt{N}(\bar{x}-\theta_0) < d_\alpha\sigma|\theta_1) \\ &= P(\sqrt{N}(\bar{x}-\theta_1) < d_\alpha\sigma - \sqrt{N}(\theta_1-\theta_0)|\theta_1)\end{aligned}$$

であるから d_β が $N(0,1)$ の上側 β 点ならば

$$-d_\alpha\sigma + \sqrt{N}(\theta_1-\theta_0) = d_\beta\sigma$$

となり，これから

$$N = \frac{\sigma^2(d_\alpha+d_\beta)^2}{(\theta_1-\theta_0)^2} \tag{7c. 4. 4}$$

が得られる．

H_0 が真のとき，標本の一定の大きさ N (7c. 4. 4) にたいする A. S. N. (7c. 4. 1) の比は

$$-\frac{2[b(1-\alpha)+a\alpha]}{(d_\alpha+d_\beta)^2} \tag{7c. 4. 5}$$

であり，これは $(\theta_1-\theta_0)$ には依存しない．同様に H_1 が真のとき，比は

$$\frac{2[b\beta+a(1-\beta)]}{(d_\alpha+d_\beta)^2} \tag{7c. 4. 6}$$

である．

$\alpha=0.05$，$\beta=0.05$ のとき，$d_\alpha=d_\beta=1.6449$ で a, b の値は次のようになる．

$$a = \log\frac{1-\beta}{\alpha} = 2.9444, \qquad b = \log\frac{\beta}{1-\alpha} = -2.9444.$$

比 (7c. 4. 5) と (7c. 4. 6) の値はそれぞれ 0. 4897 に等しく，これは，逐次的方法により観測数が 50 %程度節約されることを示している．

Chernoff (1959) は α, β が小さくて対立仮説が帰無仮説に近い一般的な場合に観測数の節約を研究した．それによると，大標本では，大きさ一定の標本の検査個数と S. P. R. T. の検査個数の期待値の比のは 4 : 1 程度である．

7c. 5 逐次解析の基本等式

補助定理 1. z は次式をみたす確率変数であるとする．

(a) $P(z>0)>0$，かつ $P(z<0)>0$

(b) $\phi(t)=E(e^{tz})$ は任意の実数値 t にたいして存在し，

(c) $E(z)\neq 0$.

このとき，$\phi(\tau)=1$ となるような $\tau\neq 0$ が存在する．$E(z)<0$ ならば $\tau>0$ で，$E(z)>0$ ならば $\tau<0$ である．

$P(z>0)>0$ だから $P(z>c)=\delta>0$ となる c が存在する．したがって

$$\begin{aligned}\phi(t) &= E(e^{tz}) = \int e^{tz}dF \\ &\geqq \int_{z>c} e^{tz}dF > e^{tc}P(z>c), \quad (t>0 \text{ のとき})\end{aligned} \tag{7c. 5. 1}$$

結果 (7c. 5. 1) から，$t\to\infty$ のとき $\phi(t)\to\infty$．同様にして $t\to-\infty$ のとき $\phi(t)\to+\infty$．さらに，$\phi(0)=1$ で，$t=0$ における $\phi(t)$ の傾き $\phi'(t)=E(ze^{tz})$ は $E(z)$ である．$E(z)<0$ ならば，$\phi(t)$ の曲線は $t=0$ で値が 1 で，傾きは下向きになり，$t\to\infty$ のとき ∞ に近づく．$\phi(t)$ は解析的であるから，t の別のある値 $\tau>0$ で再び値 1 をとらねばならない．同様に，$E(z)>0$ のとき，$\phi(\tau)=1$ であるような $\tau<0$ が存在する．

$\phi''(t)=E(z^2e^{tz})$ であるから，補助定理 1 の条件 (a) は，すべての t にたいし $\phi''(t)>0$ であることを意味しており，したがってこれはたかだか 1 つの極小値しか持てないことがわかる．それゆえ，$\phi(t)=1$ となる t の値 τ は 0 以外にただ 1 つしか存在しない．なぜなら，0 以外の τ_1, τ_2 で $\phi(t)=1$ となるならば，少なくとも 2 つの極小値があることになるからである．

補助定理 2. $x_1, x_2, \cdots$ を，仮説 H の下で確率密度が $p(\cdot|H)$ である確率変数 x の，互いに独立に同一の分布に従う観測値の系列とする．与えられた停止規則をもつ任意の逐次決定方式を考え，決定に到達するのに必要な観測値の数を n とする．$z_i=z(x_i)$ を任意の可測関数とすると，$P(n<\infty|H)=1 \Rightarrow$

$$E\{e^{tS_n}[\phi(t)]^{-n}|H\} = P(n<\infty|H_t) \qquad (7c.5.2)$$

ただし H_t の下での x の確率密度は

$$p(\cdot|H_t) = \frac{e^{tz(\cdot)}}{\phi(t)} p(\cdot|H) \qquad (7c.5.3)$$

であり，$S_n = z_1 + \cdots + z_n$，$\phi(t) = E(e^{tz}|H)$ である．

w_m を，第m段階で逐次決定方式が終了する R^m の領域とする．このとき，

$$E\{e^{tS_n}[\phi(t)]^{-n}|H\} = \sum_1^\infty \int_{w_m} [\phi(t)]^{-m} e^{tS_m} p(x_1|H)\cdots p(x_m|H) dv^{(m)}$$
$$= \sum_1^\infty \int_{w_m} p(x_1|H_t)\cdots p(x_m|H_t) dv^{(m)}$$
$$= P(n<\infty|H_t)$$

となる．

逐次解析の基本等式　　互いに独立に同一の分布に従う観測値の系列に基づく S. P. R. T. を考え，z を [(iii), **7c. 2**] で定義したように，仮説 H_0 と H_1 の下での確率密度の比の対数とする．H を $P(|z|>0|H)>0$ であるような任意の仮説とするとき，

$$E\{e^{tS_n}[\phi(t)]^{-n}|H\} = 1 \qquad (7c.5.4)$$

が成り立つ．これを逐次解析の基本等式とよぶ．

次式は容易に確かめられる．$P(|z|>0|H)>0 \Rightarrow P(|z|>0|H_t)>0$，ただし H_t は (7c. 5. 3) で定義したものである．したがって [(iii)(a), **7c. 2**] から，$P(n<\infty|H_t)=1$ となる．結果 (7c. 5. 4) は (7c. 5. 2) から出る．

7c. 5 のすべての命題の証明は，Bahadur (1958) によってなされている．

逐次決定方式の動作特性 O. C. (operating characteristic) 関数　　検査のために提出された1バッチの製品を合格にするか不合格にするか，あるいは何らかの処置をとるかどうかというような決定を下すための逐次決定方式を考えよう．逐次決定方式は，製品を1つずつとり出し，測定し，各段階で利用できる測定値に基づき抜取りを継続するか，決定を下すために停止するかの規則を適用することである．測定値の確率分布が仮説Hで規定されるとき，与えられた方式の下で1バッチの製品を合格にする確率を $\pi(H)$ であらわす．このとき π は可能な仮説の集合に属する H の関数と考えられ，それは逐次決定方式の O. C. 関数と呼ばれる．逐次決定方式の重要なクラスにたいする O. C. 関数を計算しよう．

測定値の確率分布がある選ばれた仮説 H_0 で規定されるとき，1バッチの製品を合格にする確率は $1-\alpha$ で，その確率分布が別に選ばれた仮説 H_1 に従うときには β であるという制約を設ける．この仕様は S. P. R. T. の下で H_0 が採択されるとき1バッチの製品を合格にし，H_1 が採択されるときはその1バッチを不合格にするという約束で，仮説 H_0 と H_1 の尤度比に基づく S. P. R. T. を用いることにより実現できると仮定しよう．このような方式にたいする $\pi(H)$ の近似値を，1つの観測値を得たときの仮説 H_0 と H_1 の尤度比の対数にあたる z の分布についての補助定理1と2の条件の下で，求めよう．

$\phi(t)$ を H の下での z の特性関数とすると，補助定理1により，$\phi(\tau)=1$ となる τ が存在する．この値を基本等式 (7c. 5. 4) に代入すると

$$1 = E(e^{\tau S_n}|H) = P\{S_n \leqslant b\}E\{e^{\tau S_n}|S_n \leqslant b\} + P\{S_n \geqslant a\}E\{e^{\tau S_n}|S_n \geqslant a\}$$
$$\doteq \pi(H)e^{\tau b} + [1-\pi(H)]e^{\tau a}$$

を得る．これから

$$\pi(H) \doteq \frac{1-e^{\tau a}}{e^{\tau b}-e^{\tau a}} \qquad (7c.5.5)$$

がでる．

平均検査個数 A. S. N.(Average Sample Number)　　(7c. 3. 4) のときと同様な論法を用いると，次の近似式が得られる．

$$E(S_n|H) \doteq b\pi(H) + a[1-\pi(H)]$$

しかし [(v), **7c. 2**] から

$$E(n|H)E(z|H) = E(S_n|H)$$

かつ $E(z|H) \neq 0$ だから

$$E(n|H) \doteq \frac{b\pi(H)+a[1-\pi(H)]}{E(z|H)} \qquad (7c.5.6)$$

となる．これは H_1 にたいして H_0 を検定するための S. P. R. T. に基づいた逐次決定方式を用いて，仮説 H の下で決定に到達するのに必要な平均検査個数の近似式である．

7c. 4 で構成した S. P. R. T. にたいして正規分布のときの近似的な O. C. 関数と A. S. N 関数を計算しよう．H_0 は $N(\theta_0, \sigma^2)$ で H_1 は $N(\theta_1, \sigma^2)$ であることが思い起こされる．H は θ が任意の $N(\theta, \sigma^2)$ とする．真の値が θ_0 や θ_1 とは異なる θ であるとき，対立仮説 θ_1 にたいする θ_0 の検定の性質を調べよう．さて

$$z = -\frac{1}{2\sigma^2}[(x-\theta_1)^2 - (x-\theta_0)^2]$$

$$\phi(t) = E(e^{tz}|H) = e^{q/2\sigma^2}$$

$$q = t(\theta_1-\theta_0)[2\theta-\theta_1-\theta_0+t(\theta_1-\theta_0)]$$

である．$\phi(\tau)=1$ となる $\tau \neq 0$ の値は

$$\tau(\theta) = \frac{\theta_1+\theta_0-2\theta}{\theta_1-\theta_0}$$

であり，この場合 (7c. 5. 5) を用いると O. C. 関数は

$$\pi(\theta) = \frac{1-e^{a\tau(\theta)}}{e^{b\tau(\theta)}-e^{a\tau(\theta)}}$$

で，(7c. 5. 6) を用いると A. S. N. 関数は

$$E(n|\theta) = \frac{b(1-e^{a\tau(\theta)}) + a(e^{b\tau(\theta)}-1)}{(e^{b\tau(\theta)} - e^{a\tau(\theta)})E(z|\theta)}$$

である．ただし $E(z|\theta) = [(\theta-\theta_0)^2 - (\theta-\theta_1)^2]/2\sigma^2$ である．

7c.6　逐次推定

母数の逐次推定の基本的な考え方は，予め定められた精度をもつ推定値を得るのにちょうど十分な個数を抜取るとか，あるいは，一定の抜取り費用にたいして精度を最大（すなわち危険を最小）にすることである．区間推定の場合には，精度は与えられた信頼係数をもつ区間の長さによってあらわされる．点推定では，それは平均2乗誤差かまたはある適当な損失関数となるであろう．しかし，逐次推定についてのこの主題は十分には明らかにされていない．いくつかの問題が Anscombe (1952, 1953), Wolfowitz (1947) らによって考えられてきた．本節では，分散が未知の正規分布の平均について，2段階の抜取りにより，一定の長さの区間推定値を得るための Stein (1945) の興味ある方法を考えよう．

[(v), **7b.2**] で，平均値 $\bar{x}$ と分散の不偏推定量 s^2 を与える一定の大きさ n の標本に基づいた正規分布の平均 μ の $1-\alpha$ の信頼区間推定量は

$$\bar{x} \pm \frac{st_{\alpha/2}}{\sqrt{n}} \tag{7c.6.1}$$

であったことを思い起こすだろう．ただし $t_{\alpha/2}$ は自由度が $n-1$ の t 分布の上側 $\alpha/2$ 点である．しかし信頼区間の長さ $(2st_{\alpha/2})/\sqrt{n}$ は，それ自身確率変数で，どんな特別な場合にも大きくなることがある．一方，σ^2 が未知のとき，指定された長さの信頼区間をつくることは，1つの標本からは不可能であることが知られている (Dantzig (1940))．だが，2段階の逐次抜取りを用いることにより，この問題は解決できる．

前もって定められた大きさ m の標本 $x_1, \cdots, x_m$ を抜取り，$\bar{x}, s^2$ を計算する．n は次式をみたす最小の整数 ($\geqslant m$) とする．

$$\frac{s^2}{n} \leqslant \frac{c^2}{b^2} \tag{7c.6.2}$$

ただし，b^2 は自由度 $m-1$ の t 分布の分散で c は指定された定数である．次に，さらに $n-m$ 個の観測値からなる標本を抜取る．$\bar{x} = (x_1 + \cdots + x_n)/n$ を，すべての観測値に基づいた平均値とすると，次の結果を得る．

(i)　$\bar{x}$ は，その標準偏差が $\leqslant c$ である μ の不偏推定値である．

$$\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \quad と \quad \frac{(m-1)s^2}{\sigma^2} \tag{7c.6.3}$$

が互いに独立にそれぞれ $N(0,1)$ と自由度 $m-1$ の χ^2 分布に従い，ゆえに

$$\frac{\sqrt{n}(\bar{x}-\mu)}{s} \sim t \quad (自由度\ m-1) \tag{7c.6.4}$$

$$E\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} = 0 \tag{7c.6.5}$$

であることは容易に証明できる．

n 自身が確率変数だから，(7c.6.5) から自動的に $E(\bar{x}-\mu)=0$ とはならないことに注意せよ．さて

$$\begin{aligned} E(\bar{x}-\mu) &= E\left[\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}\cdot\frac{\sigma}{\sqrt{n}}\right] \\ &= E\left[\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}\cdot f(s^2,\sigma)\right] \quad (n\ は\ s^2\ に依存するから) \\ &= E\left[\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}\right]E[f(s^2,\sigma)] \qquad (7c.6.6) \\ &= 0 \quad ((7c.6.5)\ を用いて) \end{aligned}$$

である．(7c.6.6) の段階は (7c.6.3) からわかるように $\sqrt{n}(\bar{x}-\mu)/\sigma$ と s^2 の独立性からでる．

さらに

$$\begin{aligned} E(\bar{x}-\mu)^2 &= E\left[\frac{n(\bar{x}-\mu)^2}{s^2}\cdot\frac{s^2}{n}\right] \\ &\leqslant E\left[\frac{n(\bar{x}-\mu)^2}{s^2}\cdot\frac{c^2}{b^2}\right] \\ &= \frac{c^2}{b^2}E\left[\frac{n(\bar{x}-\mu)^2}{s^2}\right] = \frac{c^2}{b^2}\cdot b^2 = c^2 \qquad (7c.6.7) \end{aligned}$$

であることがわかる．

(ii)　$2l$ を μ の信頼区間推定量のあらかじめ定めた長さとする．このとき，推定量 $\bar{x} \pm l$ の信頼係数は，$c = bl \div t_{\alpha/2}$ と選べば，$1-\alpha$ より小さくない．

$$\begin{aligned} P(|\bar{x}-\mu|<l) &= P\left(\frac{|\sqrt{n}(\bar{x}-\mu)|}{s} < \frac{\sqrt{n}\,l}{s}\right) \\ &\geqslant P\left(\frac{|\sqrt{n}(\bar{x}-\mu)|}{s} < \frac{bl}{c}\right) \qquad (7c.6.8) \end{aligned}$$

最後の確率 (7c.6.8) は $bl/c = t_{\alpha/2}$，つまり $c = bl/t_{\alpha/2}$ のとき，値が $1-\alpha$ である．

指定された長さの区間推定量を得るこの2標本方式は，このように第1段階の標本の大きさ m を任意に固定すれば完全に規定される．m の異なる選択の結果は調べることができる．第2の標本が与える σ^2 に関する情報は，この問題の解には利用されていないことに注意せよ．このことはスタインの方式の改善を可能にするであろう．

7c.7　検出力が1の逐次検定

7c.1—7c.5 で2つの対立する仮説のうち一方あるいは他方を採択して確率1で終わる

逐次検定を考えた．一方，連続生産のように，ある仕様（帰無仮説）に従って製品が生産されるかぎり何の処置も必要でないが，それ以外のとき（対立仮説が真であるとき）は処置をとる必要がある場合がある．つまり換言すれば次のような逐次決定方式を必要とする．

（a）　帰無仮説が真のときは，終了することがまれであり，

（b）　対立仮説が真のときは，できるかぎり速やかに，確率 1 で終了する．

このような逐次決定方式は最近 Barnard (1969)，ダーリングとロビンス (Robbins and Siegmund (1969) を見よ) によって，**2c.6** で述べた有名な重復対数の法則を用いて示唆されたその法則をもう一度述べよう．

重複対数の法則　$X_1, X_2, \cdots$ を $E(X_1)=\mu$, $V(X_1)=\sigma^2<\infty$ であるような互いに独立に同一の分布に従う確率変数とする．$Z_n=(X_1+\cdots+X_n-n\mu)/\sigma\sqrt{n}$ とすると

$$P\{Z_n>(1+\varepsilon)(2\log\log n)^{\frac{1}{2}} \text{ が無限に多く生起する}\}=0 \qquad (7c.7.1)$$

$$P\{Z_n>(1-\varepsilon)(2\log\log n)^{\frac{1}{2}} \text{ が無限に多く生起する}\}=1 \qquad (7c.7.2)$$

である．(7c.7.1) と (7c.7.2) の結果として

$$P(\text{ある } n\geqslant 1 \text{ にたいして } Z_n>(1-\varepsilon)(2\log\log n)^{\frac{1}{2}}\}=1 \qquad (7c.7.3)$$

$$\lim_{m\to\infty} P\{\text{すべての } n\geqslant m \text{ にたいして } Z_n<(1+\varepsilon)(2\log\log n)^{\frac{1}{2}}\}=1 \qquad (7c.7.4)$$

を得る．

$H: p\geqslant 0.01$ にたいして $H_0: p<0.01$ を検定する問題において，1964年にバーナード (Barnard (1969) を見よ) がはじめて与えた，検出力が 1 の逐次決定方式を考えよう．ただし p は生産された製品が不良である確率とする．この方式では，製品を次々にとり出し，

$$\text{ある } n\geqslant 1 \text{ にたいし，} R_n>0.01\,n+a\sqrt{n} \qquad (7c.7.5)$$

であるとき H_0 を棄却する．ここで，R_n は連続した n 個の製品中の不良品の個数で，a は適当に選んだ定数である．$H_0: p<\delta$ で $H: p\geqslant\delta$ ならば (7c.7.5) の 0.01 を δ でおきかえる．

$Z_n=(R_n-np)/\sqrt{npq}$, $q=(1-p)$ とすると (7c.7.5) は次のように書ける．

$$Z_n>\frac{(0.01-p)\sqrt{n}}{\sqrt{pq}}+\frac{a}{\sqrt{pq}}=b_n \qquad (7c.7.6)$$

(7c.7.6) で $p\geqslant 0.01$ ならば $b_n<(1-\varepsilon)(2\log\log n)^{\frac{1}{2}}$ である．したがって (7c.7.3) を用いると，確率 1 で (7c.7.5) が成り立つ．このように対立仮説が真ならばそれを採択して確率 1 で逐次決定方式は終わる．一方，$p<0.01$ ならば (7c.7.6) の b_n は $(1+\varepsilon)(2\log\log n)^{\frac{1}{2}}$ を超え，(7c.7.5) の確率は a を適当に選ぶことにより小さくできる．a が大きいほど確率は小さくなる．実際，(7c.7.5) の確率が与えられた $p=p_0<0.01$ にたいし，ある指定された値 α を持つように a を選ぶことができる．

ロビンスとダーリング (Robbins and Siegmund (1969) を見よ) による第 2 の例は $\theta>0$ にたいして $\theta\leqslant 0$ を検定する例である．ただし θ は分散が 1 の正規母集団の平均である．互いに独立な観測値 $X_1, X_2, \cdots$ が 1 つずつ得られ，観測はある $n\geqslant 1$ にたいし

$$s_n=X_1+\cdots+X_n\geqslant\sqrt{n}\,f_n \qquad (7c.7.7)$$

であるとき仮説 $\theta>0$ を採択して終わる．ここで a と m のある選ばれた値にたいして

$$f_n=\left[\frac{(n+m)}{n}\left(a^2+\log\frac{n+m}{m}\right)\right]^{\frac{1}{2}} \qquad (7c.7.8)$$

である．

$Z_n=(s_n-n\theta)/\sqrt{n}$ と書くと (7c.7.7) は次のように書ける．

$$Z_n\geqslant f_n-\sqrt{n}\,\theta=b_n \qquad (7c.7.9)$$

$\theta>0$ ならば (7c.7.9) で $b_n>(1-\varepsilon)(2\log\log n)^{\frac{1}{2}}$ となり，(7c.7.3) を用いると (7c.7.7) が確率 1 で成り立つ．このように対立仮説を検出する検出力は 1 となる．$\theta<0$ ならば，Robbins and Siegmund (1969) が示しているように

$$P\{\text{すべての } n \text{ にたいし } Z_n<f_n-\sqrt{n}\,\theta\,|\,\theta<0\} \qquad (7c.7.10)$$

$$>P\{\text{すべての } n \text{ にたいし } \quad Z_n<f_n\,|\,\theta=0\}>1-\frac{e^{-a^2/2}}{2\Phi(a)} \qquad (7c.7.11)$$

である．ただし $\Phi(a)$ は $P(X<a)$ で $X\sim N(0,1)$．たとえば $m=1$ で $a=3$ ならば

$$f_n(1+n^{-1})^{\frac{1}{2}}[9+\log(n+1)]^{\frac{1}{2}} \qquad (7c.7.12)$$

となり観測が無限に続く確率 (7c.7.10) は

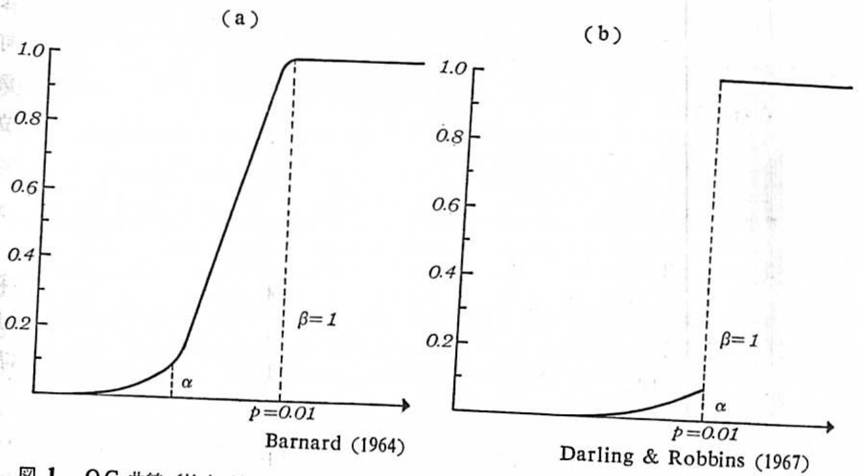

図 1.　O.C. 曲線（検定が終了する確率，母数の関数としてあらわされる）は，バーナードの方式では (a) に見られるように連続で，ダーリング＝ロビンスの方式では (b) に見られるように不連続である．第 1 種の誤りは，後者では H_0 にたいして指定された値 α よりつねに小さく，一方前者では母数の値が H_0 で選ばれた値より小さいときのみ，α より小さい．

$$1-\frac{e^{-9/2}}{2\Phi(3)}=1-0.0056=0.9944 \tag{7c.7.13}$$

より小さくない．したがって検定が終了するのは稀である．

7d　識別問題——決定理論

7d.1　問題の記述

観測された1組の特性をもとに，ある個体を有限個の群のそれが属すると思われる1つに割り当てる問題を考えよう．このような問題は，人類学者が墓地から発掘されたあごの骨からそれが男のものであるか女のものであるかを識別しようとするときや，また医者が診断のためのいくつかの検査をもとに黄疸をわずらっている患者が手術を必要とするかどうかを決定するときに起こる．また生物学者は観測した標本が可能な k 個の既知の種のどれかのメンバーであることを識別したいと思うときにも起こる．これらすべての場合において，問題はいくつかの対立する仮説のなかからの選択の問題であり，**7a** 節や **7c** 節で考えたような，ある特定の仮説を1組の対立仮説にたいして検定する問題ではない．統計学の文献では，このような問題は分類または判別問題として引用されているが，識別という言葉がよりふさわしい用語に思える．

実際の場では，生物学者が新しい種のメンバーを発見するときのように，観測した個体が，指定したどの群にも属さないことがありうると考えなければならない．このような可能性を含むような一般理論はないようである．しかしながら，第8章の **8e.2** で1つの例を考え，特殊の場合にこの問題にどのように接近できるかを示そう．本節の議論では対立仮説に関する知識は完全であると考えられている．

7d.2　確率化決定方式と非確率化決定方式

x は個体の測定値をあらわし，S を可能な x の値の（標本）空間とする．観測した x に基づいて，その個体が k 個の特定の母集団のどれに属するかについての決定を行わなければならない．この問題は，対立仮説の所与の組から，観測した事象に適切な仮説を選ぶ問題と同じである．

非確率化決定方式は空間 S を k 個の互いに排反な領域 $w_1,\cdots,w_k$ に分割し，測定値 x が $x\in w_i$ である個体を i 番目の母集団に割り当てる，つまり $x\in w_i$ なら i 番目の対立仮説を選ぶ方式である．

確率化決定方式は $\lambda_i(x)\geqslant 0$ で $\sum\lambda_i(x)=1$ であるような x のベクトル関数 $[\lambda_1(x),\cdots,\lambda_k(x)]$ を定め，測定値が x である個体を確率 $\lambda_i(x)$，$i=1,\cdots,k$ で i 番目の母集団に割り当てる方式である．すなわち x が観測された後，値 $1,\cdots,k$ をそれぞれ確率 $\lambda_1(x),\cdots,\lambda_k(x)$ でとる確率変数の観測値を生成するような確率実験が行われる．その結果が i ならば，その個体は i 番目の母集団に割り当てられる．

損失ベクトル　$P_1(x),\cdots,P_k(x)$ を k 個の母集団において σ 有限の測度 v に関する x での確率密度とし，r_{ij} を i 番目の母集団の要素を j 番目の母集団に割り当てたときの損失とする．その個体が本当は i 番目の母集団のものであるとき，与えられた方式を用いる際の期待損失は，非確率化方式にたいしては

$$L_i=\int_{w_1}r_{i1}P_i(x)dv+\cdots+\int_{w_k}r_{ik}P_i(x)dv \tag{7d.2.1}$$

であり，確率化方式にたいしては

$$L_i=\int[r_{i1}\lambda_1(x)+\cdots+r_{ik}\lambda_k(x)]P_i(x)dx \tag{7d.2.2}$$

である．どちらの場合にも k 個の対立する仮説に対応する損失ベクトル

$$(L_1,\cdots,L_k) \tag{7d.2.3}$$

を得る．これは，決定方式の動作特性として，われわれの研究では重要な役割を演ずる．

2つの決定方式 δ_1,δ_2 があり，それに関連した損失ベクトルが $(L_{11},\cdots,L_{k1}),(L_{12},\cdots,L_{k2})$ であるならば，

$$L_{i1}\leqslant L_{i2},\qquad i=1,\cdots,k \tag{7d.2.4}$$

で，少なくとも1つの i について $L_{i1}<L_{i2}$ のとき，δ_1 は明らかに δ_2 より良い．等号がすべての i について成り立つならば，2つの方式は同等である．一方，どの2つの決定方式も比較可能であるとはかぎらない．つまり，いくつかの i の値にたいしては $L_{i1}>L_{i2}$ で，残りの i についてはその逆の関係が成り立つことがある．このような方式のあいだで選択するには，さらに基準を設けなければならない．このことから許容的決定方式の概念が導かれる．

許容的方式　(7d.2.4) の意味で，より良い決定方式が他になければ，その決定方式は許容的である．許容的決定方式の集合が1つの要素から成るときは，最適解が得られる．しかし，一般にその集合は大変広く，許容的集合に属するどの2つの決定方式も比較可能ではない．許容的決定方式の集合から1つを選択するためには，さらに基準が必要である．しかしこの問題は，**7d.3** で示すように k 個の母集団の事前確率がわかっているときには簡単に解くことができる．

7d.3　ベ イ ズ 解

$\pi_1,\cdots,\pi_k$ を k 個の母集団の先験確率とする．すなわち，観測された個体は k 個の母集

団の個体が $\pi_1, \cdots, \pi_k$ なる比率で混合している母集団から選ばれたものと考える．このような場合の期待損失は

$$L = \pi_1 L_1 + \cdots + \pi_k L_k \tag{7d.3.1}$$

なる1つの量に帰着する．ここで $L_1, \cdots, L_k$ は (7d.2.3) で定義したものである．L_i に式 (7d.2.1) を用いる非確率化決定方式では (7d.3.1) は次式になる．

$$\begin{aligned} L &= \sum_{i=1}^{k} \int_{w_i} (\pi_1 r_{1i} P_1 + \cdots + \pi_k r_{ki} P_k) dv \\ &= \int_{w_1} - S_1 dv + \cdots + \int_{w_k} - S_k dv \end{aligned} \tag{7d.3.2}$$

ここで $S_i = -(\pi_1 r_{1i} P_1 + \cdots + \pi_k r_{ki} P_k)$ を (i 番目の母集団にたいする) i 番目の個体の判別得点とよぶ．確率化決定方式では次のようになる．

$$L = \int -[\sum \lambda_i(x) S_i(x)] dv \tag{7d.3.3}$$

(i) $w_1^*, \cdots, w_k^*$ は標本空間全体を覆う互いに排反な領域で，$i = 1, \cdots, k$ について

$$x \in w_i^* \Rightarrow S_i(x) \geqslant S_j(x), \qquad \text{すべての } j \text{ にたいして} \tag{7d.3.4}$$

とする．w_i をこのように選べば (7d.3.2) の期待損失 L は最小となる．

この結果は 7a.2 の補助定理4から直接得られる．

条件 (7d.3.4) には等号があるので，w_i を一意的には規定しないことが確かめられる．たとえば特定の x にたいして

$$S_{i_1}(x) = \cdots = S_{i_r}(x) > S_{i_{r+1}}(x) \geqslant \cdots \geqslant S_{i_k}(x) \tag{7d.3.5}$$

ならば，x は領域 $w_{i_1}^*, \cdots, w_{i_r}^*$ のうちのどれにおくこともできる．しかし条件 (7d.3.4) がみたされるかぎり最小損失は同じである．実際に (7d.3.5) のような場合では，確率 $\lambda_{i_j}(x)$ で x を $w_{i_j}^*$ におくように選ぶことができる．ここで $\lambda_{i_1}(x), \cdots, \lambda_{i_r}(x)$ は，その和が1という条件の下で任意に選ばれる．

(ii) $S_i(x)$ が他のどの $S_j(x)$ よりも大きいならば，$\lambda_i(x) = 1, \lambda_j(x) = 0,\ j \neq i$ とし

$$S_{i_1}(x) = \cdots = S_{i_r}(x) > S_{i_{r+1}}(x) \geqslant S_{i_{r+2}}(x) \geqslant S_{i_k}(x)$$

ならば，$\lambda_{i_1}(x), \cdots, \lambda_{i_r}(x)$ は $\lambda_{i_1}(x) + \cdots + \lambda_{i_r}(x) = 1,\ \lambda_{i_{r+1}}(x) = \cdots = \lambda_{i_k}(x) = 0$ なる条件の下で任意であるとしよう．このように $\lambda_i(x)$ を選べば (7d.3.3) の損失 L は最小となる．

この結果は 7a.2 の補助定理3から直接得られる．

確率化方式にたいする最適解と非確率化方式にたいする最適解とは本質的に同じであることが確かめられる．(ii) で規定される決定方式を，事前分布または事前確率ベクトル $\boldsymbol{\pi}' = (\pi_1, \cdots, \pi_k)$ に関するベイズ方式とよぼう．対応する損失ベクトルを $\mathbf{L}_\pi' = (L_{1\pi}, \cdots, L_{k\pi})$ であらわす．するとベイズ解は

$$\boldsymbol{\pi}' \mathbf{L}_\pi \leqslant \boldsymbol{\pi}' \mathbf{L} \tag{7d.3.6}$$

なる性質により特徴づけられる．ただし $\mathbf{L}$ は任意の他の (確率化または非確率化) 決定方式にたいする損失ベクトルである．

ベイズ解はまた，別の種類の論法で導くことができる．測定値が x である個体が i 番目の母集団に属する条件つき確率，つまり x が観測されたとき，i 番目の母集団の事後確率は，ベイズの定理により

$$\frac{\pi_i P_i}{\sum \pi_i P_i}, \qquad i = 1, \cdots, k \tag{7d.3.7}$$

となる．測定値が x である個体を j 番目の母集団に割り当てると決めるならば，条件つき損失は

$$\frac{r_{1j}\pi_1 P_1 + \cdots + r_{kj}\pi_k P_k}{\sum \pi_i P_i} \tag{7d.3.8}$$

である．これは $-S_j(x)/\sum \pi_i P_i$ に等しい．ただし S_i は (7d.3.2) で定義したものである．与えられた x にたいし，(7d.3.8) が最小つまり判別得点が最高となるような母集団にその個体を割り当てたとき，条件つき損失は最小となる．このような決定方式にたいして条件つき損失が最小であるから，全期待損失は明らかに最小である．この決定方式は，実は，$\boldsymbol{\pi}$ に関するベイズ方式と同一である．

$r_{ij} = 1,\ i \neq j$ かつ $r_{ii} = 0$ のとき，L_i は i 番目の母集団の個体にたいする誤った識別の期待比率をあらわし，L は事前確率ベクトル $\boldsymbol{\pi}$ を用いたときのすべての個体にたいするそれをあらわす．$r_{ij} = 1,\ i \neq j$ かつ $r_{ii} = 0$ のとき i 番目の母集団にたいする判別得点は

$$S_i = -\sum_1^k \pi_j P_j + \pi_i P_i = \pi_i P_i + c \tag{7d.3.9}$$

である．ここで c は i によらない定数である．このとき，測定値が x である個体は次式を満足する1つの母集団に割り当てられる．

$$\text{すべての } j \text{ にたいし，} \pi_i P_i(x) + c \geqslant \pi_j P_j(x) + c$$

つまり，すべての j にたいし

$$\pi_i P_i(x) \geqslant \pi_j P_j(x). \tag{7d.3.10}$$

方式 (7d.3.10) は誤った識別の期待回数を最小にする．この目的のために S_i，つまり i 番目の判別得点を単に $\pi_i P_i$ と定義できる．

7d.4　決定方式の完全類

多くの実際の問題では，事前確率はわからないか，または不適切であるかもしれない．そのような場合，すべての可能な決定方式から1つを選ぶのに何らかの基準が必要である．

このようなことをもくろんで，まず選択をしなければならない方式の，最小の集合に到達できるような方法を調べてみよう．

決定方式の集合Cを，Cに属しない任意の決定方式δにたいしてδより良いCの要素が存在するとき，完全類という．

ある完全類を部分集合として含まない完全類を最小完全類という．次に決定方式の完全類および最小完全類を特徴づけるいくつかの命題を証明しよう．

（i）最小完全類が存在するための必要十分条件は，許容的決定方式の集合が完全類であることである．

証明は容易である．

（ii）あらゆる決定関数に対応するすべての損失ベクトルの集合$\mathscr{L}$はE_k（k次元ユークリッド空間）の閉凸集合である．

証明は省略する．Wald (1950) を参照するとよい．

（iii）あらゆる許容的決定方式の集合Aは最小完全類である．

Aが完全であることさえ証明すればよい．方式$\delta_1 \notin A$を考えよう．これは明らかに許容的ではない．Δをδ_1より良い決定方式の集合とし，δに対応する損失ベクトルの成分の和を$f(\delta)$であらわす．

$$f(\delta_j) \leqslant \inf_{\delta \in \Delta} f(\delta) + \frac{1}{j}$$

となるようなδ_jを選ぶ．$\mathscr{L}$は(ii)により閉じているから，jの値のある部分列が存在して，$\lim \delta_j = \delta^*$が存在し，かつそれはδ_1より良いようにできる．さらに任意の方式$\delta_0 \in \Delta$にたいして

$$f(\delta_0) \geqslant \inf_{\delta \in \Delta} f(\delta) \geqslant f(\delta_j) - \frac{1}{j} \to f(\delta^*)$$

となるから，δ^*は許容的である．$\delta^* \in A$だからこのようにしてAは完全類となる．

（iv）どの許容的方式も，ある事前分布に関するベイズ方式である．

$\mathbf{L}_0$を方式$a_0 \in A$に対応する損失ベクトルとする．$\mathscr{L}$で$\mathbf{L}$を変え，$\mathbf{L}_0$を固定するとき，ベクトル$\mathbf{L}-\mathbf{L}_0$の全体は凸集合になり，a_0が許容的だから，そのどの要素もすべての成分が負であることはない．したがって，支持平面定理[(iv), **1d. 2**]を適用すると，すべての$\mathbf{L} \in \mathscr{L}$にたいし

$$\boldsymbol{\pi}'(\mathbf{L}-\mathbf{L}_0) \geqslant 0 \tag{7d. 4. 1}$$

となるような非負の成分をもつベクトル$\boldsymbol{\pi}$が存在する．この結果 (7d. 4. 1) は，$\mathbf{L}_0$が$\boldsymbol{\pi}$に関するベイズ方式にたいする損失ベクトルであることを示している．

（v）すべてのベイズ方式の集合$\mathscr{B}$は完全類である．

この結果は (iii) から導かれる．

（vi）任意の$\boldsymbol{\pi}$に関するベイズ方式は，$\boldsymbol{\pi}$のすべての成分が正ならば許容的である．

$\mathbf{L}_\pi$を，$\boldsymbol{\pi}$に関するベイズ方式に対応する損失ベクトルとすると

$$\boldsymbol{\pi}'\mathbf{L}_\pi \leqslant \boldsymbol{\pi}'\mathbf{L}, \qquad \mathbf{L} \in \mathscr{L}$$

である．

$\mathbf{L}_0 \in \mathscr{L}$が存在し，$\mathbf{L}_0$が$\mathbf{L}_\pi$よりも良いならば

$$\boldsymbol{\pi}'\mathbf{L}_0 \leqslant \boldsymbol{\pi}'\mathbf{L}_\pi \leqslant \boldsymbol{\pi}'\mathbf{L}_0$$

を得る．$\boldsymbol{\pi}$のすべての成分は正だから$\mathbf{L}_\pi = \mathbf{L}_0$である．

7d. 5　ミニマックス方式

7d. 4節では，許容的方式の集合は大変広く，決定方式を選ぶにはさらに基準を設ける必要があることを示した．このような原理の1つは，ミニマックスとして知られ，それは損失ベクトルの最大の成分が最小値をもつような決定方式を選ぶことである．つまりその方式δ^*の損失ベクトルを$(L_1^*, \cdots, L_k^*)$とするとき，それがもし存在するならば，次式をみたす．

$$\max_i (L_1^*, \cdots, L_k^*) = \min_\delta \max_i (L_1 \cdots, L_k) \tag{7d. 5. 1}$$

ここでiは指数$1, \cdots, k$をとり，δは許容的決定方式の集合上を動く．このような方式を求めることは，たとえそれが存在するとしても，困難であろう．しかし，決定方式がミニマックス方式と認められるような場合は存在する．

（i）正の確率ベクトル$\bar{\boldsymbol{\pi}}' = (\bar{\pi}_1, \cdots, \bar{\pi}_k)$が存在して，これに対応するベイズ解の損失ベクトルの成分がすべて等しいとする．このとき$\bar{\boldsymbol{\pi}}$を最も不利な事前分布とよぶと，$\bar{\boldsymbol{\pi}}$に対応するベイズ決定方式はミニマックス方式である．

定義により，このような方式の損失ベクトルは

$$(\bar{L}, \cdots, \bar{L}) \tag{7d. 5. 2}$$

であり，正の事前確率をもつベイズ決定方式は，すべて許容的であるから，最大成分が$\bar{L}$より大きくない損失ベクトルは存在しない．もしそうでなければ (7d. 5. 2) は許容的でない．

ミニマックス解の存在の証明や **7d. 4** と **7d. 5** の結果の可算個，あるいは非可算無限個の対立仮説への拡張については，読者は Blackwell and Girshick (1954) や Wald (1950) による本を参照せよ．**7d** で論じた識別の問題は一般的な決定理論，あるいはより一般的なゲーム理論の特定の場合である．この問題にたいする良い，初等的な入門書として，読者は Chernoff and Moses (1959) による決定理論の本を参照されたい．

7e　ノンパラメトリック推測

7e.1　頑健性の概念

これまでの各章で論じた統計的方法の目的は以下の通りであった．どの問題でも，われわれは一組の観測値と，その観測値の確率分布に関する部分的な情報を得る．与えられた情報は正しいと仮定して，データを解析したり，確率分布の未知の側面（つまり特徴）に関して推測したりするための最適方式が求められた．したがって，厳密にいえば，ある特定の方法の有効性は，与えられた情報の正確さに依存している．観測値が得られた母集団の未知の平均についての推測を行うために実際に広く用いられているスチューデントの t 統計量に関して，この状態を復習してみよう．ここで，

$$t = \frac{\sqrt{n}\,|\bar{x} - \mu_0|}{s} \tag{7e.1.1}$$

（ただし $\bar{x}$ と s は確率変数 X の n 個の観測値 $x_1, \cdots, x_n$ から計算される）で定義される t 統計量の分布は，次の仮定の下で得られる．

（a）　確率変数 X の分布は正規分布である．

（b）　得られた観測値は互いに独立である．

（c）　観測値を記録する際には誤差がない．

（e）　母集団の平均は正確に μ_0 である．

与えられた母集団の個体の平均身長が $\mu_0 = 70$ インチであるという仮説を，n 個の観測値に基づいて検定するために，t 分布を用いたいとしよう．選ばれたある（小さな）水準で有意となるような大きな t の値が得られたとき，いかなる推測が行えるであろうか？(a) から (d) までの仮定の1つ以上が間違っていたのかもしれないし，もしそうなら，すべての仮定が正しいにもかかわらずまれな事象が起きたという可能性は考えなくてもよいことになる．1つずつ調べてみよう．

母集団の個体の数は，大きいけれども有限である．したがって個体の身長のような特性は，有限個の異なる値をとるにすぎない．だから，個体の身長にたいして，正規分布のような連続分布を仮定することは厳密には正しくない．しかし，一部は理論的な，また一部は経験的な研究から，正規性からのあまり大きくないずれにたいして，t 分布は敏感ではなく，したがってその適用は，正規性の仮定(a)によって厳しく制限されることはないことが知られている．このような性質は頑健性 (Box and Anderson (1955)) といわれている．ゆえに有意な t は，観測値の正規性からのずれを示すものとは考えられないであろう．

観測値の非独立性の，t 分布への影響はどうであろうか？　すべての観測値は互いに相関があり，任意の2つのあいだの相関は共通の ρ であるとしよう．すると

$$V(\bar{x}) = \frac{\sigma^2}{n}(1 + \overline{n-1}\rho) = E(\bar{x} - \mu_0)^2 \tag{7e.1.2}$$

$$E(s^2) = \sigma^2(1 - \rho) \tag{7e.1.3}$$

となる．t 統計量 (7e.1.1) の代わりに

$$t^2 = n(\bar{x} - \mu_0)^2 \div s^2 \tag{7e.1.4}$$

を考えよう．これは自由度が $1, n-1$ の F 分布に従う．(7e.1.2) と (7e.1.3) から t^2 の分子と分母の期待値は

$$\sigma^2(1 + \overline{n-1}\rho) \quad と \quad \sigma^2(1 - \rho) \tag{7e.1.5}$$

となる．期待値の比は $\rho = 0$ のとき1だが，$\rho > 0$ のとき1より大きく，$\rho \to 1$ のとき $\to \infty$ になる．このように ρ が正で大きいとき，たとえ μ_0 が真の平均値だとしても，t の大きな値が起こることが予想される．したがって有意な t は仮定 (b) からのずれによるものであるかもしれない．

記録誤差の t 分布への影響を調べる方法はないが，少し注意すれば仮定 (c) からのずれは避けられる．

最後に，仮定 (a) から (c) が正しくて，真の平均値が $\mu \neq \mu_0$ であるとき，(7e.1.4) の t^2 の分子と分母の期待値の比は

$$\frac{n(\mu - \mu_0)^2}{\sigma^2} + 1 \tag{7e.1.6}$$

であり，$\mu = \mu_0$ のとき1になる．したがって $|t|$ の大きな値は仮定 (d) が誤りのときに起こる．これがまさに，分布の平均に関する帰無仮説を検定するのに，t 検定が使われる所以である．

式 (7e.1.5) の計算では，共通の相関 ρ をもつという相互従属の極端な場合を考えた．しかし，一般に何対かの変数に正の相関を与えるような従属性があると，t の有意性を過大評価すると考えられている．個体を選択するのに乱数を使用するような無作為化操作を採用することにより，観測値間に相関がもちこまれていないと確信できる場合には，t の有意性は指定された平均値からのずれを示すものだと解釈してよい．

他の検定基準についての同様な吟味は，この章の終わりの問題 **1, 2, 3** を見よ．分散分析の手法に基づく検定基準の吟味については，Plackett (1960) と Scheffe (1959) を見よ．

7e.2　分布によらない方法

7e.1 で，統計的手法が非常に異なる条件の下で演ずる役割やその動作特性を調べることの重要性がわかった．このことは，任意の統計的手法が与えると期待される新しい知識

の性質や，特定の推測を有効にするために要求される前提条件を吟味する際に必要となる．もし統計的手法が，はじめから是認しなければならない仮定とか，データ自身によっては確かめることができない仮定とか，あるいは，なにか他の論理的根拠では正当化されないような仮定とかに依存するとすれば，それは新しい知識をつくり出す道具とはならないであろう．したがって統計的な道具は，特定のモデルにたいして適切であるのではなく，いろいろなモデルを通じて実験データを解釈するときの助けとなるものとして，よりふさわしいと考えるべきである．

有意な t が得られたとき，“もし **7e.1** の仮定 (a) から (c) までがみたされるならば，指定された平均値は多分間違いであろう” と述べるのは，有用な叙述とはならないであろう．この原理に動機づけられて，統計学者はできるだけ少ない仮定に依存する統計的手法をつくり出そうと試みた．このようにして発展したのがノンパラメトリック法として知られる手法である——これは，観測値が得られた確率変数の分布に依存しないで適用できるので，分布によらない方法とよぶ方がより適切である．

このような検定法はあまり敏感でないかもしれない．一方，多くの実際面では，観測値の変換をすることによって近似的に正規性の仮定が満足されるようにできるかもしれない．そのような場合には，正規理論に基づく最適検定手法の多くが適用可能となる．いくつかのノンパラメトリック検定を簡潔に考察し，その多様性を示そう．

7e.3　いくつかのノンパラメトリック検定

符号検定　$x_1, \cdots, x_n$ が，連続分布関数をもつ確率変数 X について互いに独立に同一の分布に従う観測値であるとする．X のメディアンが μ_0（所与の値）であるという仮説を考えよう．

$(x_1-\mu_0), \cdots, (x_n-\mu_0)$ のうちで負であるものの数を r，正であるものの数を s とする．ただし $r+s \leqslant n$ である．読者は，$r+s$ を与えたときの r の条件つき分布が，成功の確率 π が 1/2 に等しい2項分布であることを証明できるであろう．当面の仮説は，$r+s$ 回のベルヌイ試行のうち r 回の成功が起こったという条件の下で，$\pi=1/2$ なる仮説を検定することと同等である．これにたいしては2項確率の表に基づく正確な検定か，あるいは，$r+s$ が大きいときは自由度1のカイ2乗検定を用いることができる．

2標本検定　$x_1, \cdots, x_m$ を，連続分布関数 F_1 をもつ X の互いに独立に同一の分布に従う観測値とし，同様に $y_1, \cdots, y_n$ を連続分布関数 F_2 をもつ Y の観測値とする．仮説 $F_1=F_2$ を検定するのに，いくつかのノンパラメトリック検定がある．この目的のためのカイ2乗検定やコルモゴロフ＝スミルノフ検定は，すでに **6b, 6c, 6f** 節で述べた．それらは，帰無仮説 $F_1=F_2$ からの種々様々なずれを検出するには有用である．しかし主要な関心が

平均値の差のみを検出することにあるならば，次のいくつかの検定が役に立つであろう．

ワルド＝ウォルフォヴィッツ (1940) の連の検定　2標本を併合した $(m+n)$ 個の観測値を，まずその大きさの上昇順に並べる．第1の標本の各観測値を0，第2の標本のそれを1でおきかえると，0と1の系列ができる．$F_1=F_2$ ならば，m 個の0と n 個の1のすべての可能な $[(n+m)!/(n!m!)]$ 個の配列は同等に起こりやすいから，0と1の系列に基づく任意の統計量の分布は正確に求めることができる．その1つの統計量は0と1の連の数である．$F_1 \neq F_2$ で，特に F_1 が F_2 と分布の位置の移動だけ異なるときは，連の期待個数は小さくなるであろうから，観測された連の数が，選ばれた有意水準で指定された数より小さいならば，有意であると宣言できる．

ウィルコクソン (1945) 検定　$(m+n)$ 個の観測値をワルド＝ウォルフォヴィッツ検定のときのようにその大きさの上昇順に並べると，第2の標本の観測値の順位が得られる．その順位を $s_1, \cdots, s_n$ とする．検定基準は順位の和である．帰無仮説が真ならば，$1, 2, \cdots, m+n$ から n 個の順位（数）をどのように選んでもそれらは同じ確率 $n!m!/(n+m)!$ をもつから，順位の和の分布を求めることができる．片側対立仮説を考えると，その統計量（順位の和）の大きな値は帰無仮説からのずれを示すであろう．対立仮説が両側ならば，統計量の小さい値と，大きい値の両方を棄却域を規定するときに考慮に入れねばならない．

フィッシャー＝イェーツ検定　$s_1, \cdots, s_n$ を第2の標本の順位の値とする．検定基準は $\sum c(s_i)$ で，ここで $c(s_i)$ は $N(0,1)$ からの，大きさ $(m+n)$ の標本の s_i 番目の順序統計量の期待値である．$N(0,1)$ からの標本の大きさが20までの順序統計量の期待値は R. M. M. 表に，大きさ50までは Biometrika Tables, vol.1 に与えられている．片側検定と両側検定は，ウィルコクソン検定の場合のように統計量 $\sum c(s_i)$ に基づいて構成される．

ファンデルヴェルデン検定　検定基準は $\sum \Phi^{-1}[s_i/(m+n+1)]$ である．ここで Φ は標準正規確率変数の累積分布である．この検定を容易にする表は van der Waerden and Nievergelt (1956) により与えられている．

これらは現在利用できるほんの2,3のノンパラメトリック検定である．ノンパラメトリック検定の応用と，いくつかの検定統計量の有意になる値は Siegel (1956) の本に収録されている．ノンパラメトリック検定と推測の理論の一部は Fraser (1957), Puri and Sen (1970), Lehman (1959) の本に見られる．ノンパラメトリック統計学とそれに関連した話題の文献は，Savage (1953) と Walsh (1962) に集められている．

7e.4　無作為化の原理

2つの処理 A と B が個体の母集団に与える効果の差を検定したいとする．m 個の個体よ

りなる無作為標本をとり，処理Aの効果を測り，

$$x_1, \cdots, x_m \tag{7e. 4. 1}$$

を得る．これはAの個体に与える効果をあらわす確率変数Xについての1組の独立な観測値を構成する．同様にして別のn個の固体よりなる無作為標本をとり，Bの効果を測り，

$$y_1, \cdots, y_n \tag{7e. 4. 2}$$

を得る．これはBの効果をあらわす確率変数Yについての観測値である．F_1とF_2でXとYの分布関数をあらわすと，帰無仮説はF_1とF_2が等しいことと解釈される．このような場合，観測値（7e. 4. 1）と（7e. 4. 2）に **7e. 3** のノンパラメトリック検定のどれかを適用して，検討中の個体の母集団に関して，AとBの差の有効な推測を行うことができる．

しかし，一般に，少なくとも調査の予備段階では，最終の勧告をしなければならない母集団からの個体の標本を含むような実験を行うことは望ましくない．実験室的規模の実験を，志願者またはある特質に基づいて特に選ばれた人，あるいはその処理に反応する実験動物をさえ用いて行うのがよい．そのような実験の結果は，母集団全体には適用できないが，母集団の個体に関する適当な実験を計画し検討するための決定に論拠を与えることができるであろう．

次に，利用できる被験者を用いて実験し，処理の差に関する情報をもたらすような観測値を生成する問題がある．統計的推測は観測値を支配する確率模型について利用できる知識に基づくのであるから，種々の結果に関連する確率測度を帰無仮説の有意性検定を実行するのに最小限必要な程度に規定できるように，観測値を生成する必要がある．これはまさに無作為化の原理によって達成されるのであり，この原理はフィッシャーの顕著な貢献によるもので，現在広く探求されている実験計画法の分野の基礎を形成したものである．

無作為化の原理によれば，被験者への処理の割り当ては確率実験の結果に基づく．AとBの2処理があるときは，個々の場合ごとに銅貨投げにより（表がでればAを，そうでなければBを与える）AかBかを決定する．または，利用できる被験者を乱数を用いてランダムに（選ばれた大きさの）2つの群に分割し，一方の群のメンバーにAを，他方の群のメンバーにBを与える．被験者をAにm人，Bにn人割り当てるとすると，全個数$N = m + n$の個体を大きさmとnの2つの群に分割するのに$(m+n)!/m!n!$通りある．無作為化手順を用いて，そのような分割のおのおのが，実験に選びだされる機会を等しくする．このような手順は，実験の結果の標本空間にどのような確率測度を規定するであろうか？

Aの効果とBの効果は個体ごとには異なる（自然変動）かもしれないが，任意の所与の個体では同じであるということを帰無仮説とする．この帰無仮説の下でN個の個体に与えるAまたはBの効果を$z_1, \cdots, z_N$であらわす．この個体が大きさmとnの2つの群に分割され，第1の群はAを，第2の群はBを受けたとすると，Aの観測された効果は$z_1, \cdots, z_N$の部分集合である，

$$x_1, \cdots, x_m \tag{7e. 4. 3}$$

となり，このとき，その補集合

$$y_1, \cdots, y_n \tag{7e. 4. 4}$$

はBの観測された効果となるであろう．したがって，1つの実験で観測値（7e. 4. 3），（7e. 4. 4）が与えられることになり，それらを合わせると$z_1, \cdots, z_N$の組と同じになる．m個の個体よりなるどの部分集合もAに割り当てられる確率は同じであるから，$z_1, \cdots, z_N$のうちのm個の値よりなるどの部分集合もAの効果として観測される確率は同じで$n!m!/N!$である．（7e. 4. 3）と（7e. 4. 4）の平均値の差である統計量

$$d = |\bar{x} - \bar{y}| \tag{7e. 4. 5}$$

を定義すると，その分布は，$z_1, \cdots, z_N$（その値は特定の実験ではわかっている）の与えられた大きさの2群へのすべての分割を考えることによって計算できる．このような分布は並べかえによって生ずる分布として知られ，その妥当性は，個体を処理に割り当てる際に用いた手順の完全さのみに依存する．並べかえによって生ずるdの分布から，$P(d \geqslant d_\alpha) \leqslant \alpha$（$\alpha$は選ばれた有意水準）となるような値$d_\alpha$が得られる．観測された$d$が，$d \geqslant d_\alpha$のとき，帰無仮説は棄却される．

実際には，考えるべき分割の数が大きいので，Nが小さくなければ，dの正確な分布を求めることは困難であろう．しかしmとnが大きいとき，次の統計量

$$t = \sqrt{\frac{mn}{m+n}}\frac{\bar{x} - \bar{y}}{s} \tag{7e. 4. 6}$$

$$s^2 = \frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{m+n-2}$$

は，自由度$m + n - 2$のスチューデントのtとして用いることができる．

有意なtまたはdは，どのように解釈すべきであろうか．処理AとBが任意の個体に与える効果は同じであるという帰無仮説は棄却されたままである．しかし，それぞれの個体ではAとBに差があるが，全体の（つまり平均の）差は0であるという可能性があるかもしれない．したがって帰無仮説からのどのような特殊のずれをその検定は検出できるのかを調べることが必要となる．この問題に答えるには，帰無仮説が真でないときの検定手順のO. C.（動作特性）を調べればよい．

そこでもっとも一般的な場合を考えて，任意の数

$$z_{1A}, \cdots, z_{NA}$$

$$z_{1B}, \cdots, z_{NB} \tag{7e. 4. 7}$$

で，N個の個体に与えるAとBの効果をあらわすことにしよう．次式を定義する．

$$\frac{1}{N}\sum_{i=1}^{N} z_{ir} = \bar{z}_r, \qquad r = A, B$$
$$\bar{z}_A - \bar{z}_B = \Delta$$
$$\sum (z_{ir} - \bar{z}_r)^2 = (N-1)V_r, \qquad r = A, B$$
$$\sum (z_{iA} - \bar{z}_A)(z_{iB} - \bar{z}_B) = (N-1)C_{AB}.$$

A と B の観測された効果を $x_1, \cdots, x_m$ と $y_1, \cdots, y_n$ とする．一般的な模型 (7e. 4. 7) の下で，並べかえによって生ずる分布の x_i, y_i の期待値，分散，共分散は次式で与えられる．

$$E(x_i) = \bar{z}_A, \qquad E(y_i) = \bar{z}_B$$
$$V(x_i) = \frac{(N-1)V_A}{N}, \qquad \mathrm{cov}(x_i, x_j) = -\frac{V_A}{N} \tag{7e. 4. 8}$$
$$V(y_i) = \frac{(N-1)V_B}{N}, \qquad \mathrm{cov}(y_i, y_j) = -\frac{V_B}{N}$$
$$\mathrm{cov}(x_i, y_j) = -\frac{C_{AB}}{N}.$$

(7e. 4. 6) で定義した t のかわりに，統計量 t^2 を考えよう．すると

$$t^2 = \frac{mn}{m+n}(\bar{x} - \bar{y}) \div s^2 = a^2 \div b^2 \text{（とおく）}. \tag{7e. 4. 9}$$

式 (7e. 4. 8) を用いると，簡単な計算により

$$E(a^2) = \frac{nm}{m+n}\left[\Delta^2 - \frac{1}{N}(V_A + V_B - 2C_{AB})\right] + \frac{nV_A + mV_B}{m+n} \tag{7e. 4. 10}$$
$$E(b^2) = \frac{(m-1)V_A + (n-1)V_B}{m+n-2} \tag{7e. 4. 11}$$

を証明することができる．(7e. 4. 10) の第2項は，特に $m = n$ のときは (7e. 4. 11) と同じオーダーの大きさである．$\Delta = 0$ ならば $V_A + V_B - 2C_{AB} \geqq 0$ だから，t^2 の分子の期待値は分母の期待値よりもほんとうに小さくなる傾向にある．したがって，実際に処理効果は各個体にたいしては異なるが，平均の差 $\Delta = 0$ というようなことがあっても，有意に大きな t は起こりえないであろう．$\Delta \neq 0$ ならば，t^2 の分子はより大きな値になると期待されるから，有意な t は主に $\Delta \neq 0$ による，つまり効果 (7e. 4. 7) の他の面よりもむしろ平均における真の差による，と主張するのが妥当のように思われる．

実験データから有効な推測を引きだすために，無作為化の原理が活用されるいろいろな方法について，読者は Cochran and Cox (1957), Fisher (1935) や Kempthorne (1952) による実験計画法の本を参照せよ．並べかえによって生ずる分布と，ノンパラメトリック推測に関して興味あるいくつかの論文は Chernoff and Savage (1958), Hoeffding (1951), Mitra (1955), Neyman (1935), Pitman (1937, 1938), Rao (1959b) Wolfowitz (1949) によるものである．

7f　補　助　情　報

6d. 3 で，2×2 分割表における属性の独立性の条件つき検定を導入した．そこでは検定統計量の有意性は，標本で観測された周辺和を与えたときの条件つき分布を用いて判断された．Barnard (1945) はそれにかわる検定が存在し，それは検出力がより高いことを示した．しかしこの提案は，いわゆる補助情報に基づく条件つき検定を主唱した Fisher (1956) には気に入られなかった．このような条件つき検定の正確な基本方式は，まだ十分には解明されていない．しかし，フィッシャーが列挙した原理を説明しよう．興味をもった読者は，Barnard (1945), Barnard, Jenkins and Winsten (1962), Basu (1964), Birnbaum (1962), Cox (1958), Fisher (1956) の諸論文を参照せよ．

定義　$(\mathcal{X}, \mathcal{B})$ 上で定義された確率測度族 P_θ と，$\mathcal{X}$ で定義された統計量 T により誘導された分布族 P_θ^T を考えよう．P_θ^T が θ によらないとき，しかもそのときにかぎり，T を補助統計量という．

フィッシャーは，推測のためには，まず出発点として，標本における補助統計量の観測値 $T = t$ を与えたときの，$\mathcal{X}$ 上の条件つき確率測度族を考えるべきである，と提案した．それは，他のどの補助統計量 T も T_* の関数であるという意味で，いわゆる最大補助統計量とよばれる T_* が存在するならば結構であろう．この場合 T_* に条件をつけることができるであろう．不幸にも，最大補助統計量は Basu (1964) が指摘したように，存在しないかもしれない．しかし，補助情報の重要性を強調するためにいくつかの例を考えよう．5個の球を含む壺があって，そのうち未知の数 α だけが赤でそれ以外は白であるとしよう．実験では，まず偏りのない銅貨を投げて，銅貨の表がでれば1つずつ球をとりだし，毎回壺に戻しながら5回観測を行う．裏がでれば，壺のなかの5個の球全部を観測する．

このような場合，α の最尤推定値は，赤球の観測された数 r で，r の標準誤差は，$\sqrt{\alpha(5-\alpha)}/5\sqrt{10}$ となる．そのような推測，つまり最尤推定値とその標準誤差へのデータの要約は，補助情報を認めない．

表または裏の事象が補助情報を与えることを確かめよう．その場合 α の推定量は次のようになり，その標準誤差は補助統計量に依存する条件つき分布に基づく．

推定量	標準誤差	補助情報
r	$\sqrt{\alpha(5-\alpha)}/5\sqrt{5}$	表
r	0	裏

補助情報を認知しないために，推定値に誤差がないのに，強いて標準誤差を付けるとい

うような奇妙な状態に導かれたようである！　それゆえ，補助情報で補わないで最尤推定量へデータを要約すると，この例のように変則的な状態に終わることがある．

さらに一般的に実験，$E_1, \cdots, E_k$ の混合を考えよう．ただし E_i は既知の確率 π_i で選ばれるが，各実験は同じ未知母数 θ に関する情報を含んでいるとする．対立仮説 $H: \theta = \theta_1$ にたいして帰無仮説 $H_0: \theta = \theta_0$ を検定することが問題であれば，ネイマン＝ピアソンの補助定理に従う最適手順は，

$$P_1(x|E_i)/P_0(x|E_i) \geqslant k \tag{7f.1}$$

のとき H_0 を棄却することである．ここで $P_j(x|E_i)$ は，E_i が実行され H_j が真のときの，観測値 x における確率密度である．(7f.1) の k は

$$\sum_{i=1}^{k} \pi_i \,\mathrm{Prob.}\{P_1(x|E_i)/P_0(x|E_i) \geqslant k | E_i\} = \alpha \quad (\text{与えられた値}) \tag{7f.2}$$

となるように定められる．しかし，補助情報を認めると，E_i が選ばれたとき，推測は実験の他の可能性を無視して，E_i を与えたときの条件つき確率にのみ基づくべきことになる．したがって，E_i が選ばれたときの検定統計量は

$$P_1(x|E_i)/P_0(x|E_i) \geqslant k_i \tag{7f.3}$$

となり，k_i は

$$\mathrm{Prob.}\{P_1(X|E_i)/P_0(x|E_i) \geqslant k_i | E_i\} = \alpha \quad (\text{与えられた値}) \tag{7f.4}$$

となるように定められる．

検定手順 (7f.3) は，長い目で見れば，(7f.1) より低い検出力をもつ．どちらの手順を選ぶべきだろうか？　著者は後者を選びたいと思うが，熟考し決定することは読者に任せる．

補 足 と 問 題

1　<u>t 検定についての正規性の仮定からのずれ</u>　大きさ n の標本に基づく t 統計量は $\sqrt{n}(\bar{x}-\mu)/s$ である．ここで $\bar{x}$ は平均値，s^2 は補正済平方和を $n-1$ で割った値である．

1.1　もとの母集団が指数分布，$\theta^{-1}\exp(-x/\theta)$，$x \geqslant 0$ に従うとき，標本の大きさ $n=2$ にたいする t の分布を求めよ．正規理論に基づく $|t|$ 分布の5％点を用いて第1種の誤りの実際の確率を計算し，t の分布の2つのすその間を不均等に分割していることを示せ．[指数分布にたいして，歪みと尖りの尺度，$\sqrt{\beta_1}$ と β_2 の値は2と9で，正規分布にたいする値，0と3とは非常に異なることに注意せよ．]

1.2　長方形分布の母集団から大きさ5の標本を抜きとり，それぞれの場合の t 統計量を計算せよ．観測した t の値の度数分布を求め，正規理論から導かれる自由度4の t 分布と比較せよ．観測した度数分布から，正規理論に基づく t の5％と1％の値に対応する第1種の誤りの実際の確率を推定せよ．[適当な比較をするには200から500の標本が必要

である．これはクラスの学生に共同課題として与えることのできる問題である．]

2　<u>t 検定についての独立性の仮定からのずれ</u>　$x_1, \cdots, x_m$ を $E(x_i)=0$, $V(x_i)=\sigma^2$ で $|i-j|=1$ のとき $\mathrm{cov}(x_i, x_j) = \rho\sigma^2$ で，その他の場合は0であるような観測値とする．

2.1　t 統計量 $\sqrt{n}\,\bar{x}/s$ の漸近分布は，$N[0, 1+2\rho]$ であり，したがって大標本では t の有意性は $\rho>0$ のとき過大評価され，$\rho<0$ のとき過小評価されることを示せ．有限の n について，$t^2 = n\bar{x}^2/s^2$ の分子と分母の期待値を計算し，t の有意な値を仮説 $E(x_i)=0$ に矛盾するものであると解釈する際に $\rho \neq 0$ の影響を調べよ．

2.2　n が偶数のとき，

$$y_1 = \frac{x_1+x_2}{2}, \quad \cdots, \quad y_{n/2} = \frac{x_{n-1}+x_n}{2}$$

なる観測値を考え，y の標本に t 検定（自由度 $(n/2)-1$）を適用することにより，x_i の期待値について妥当な推測が行えることを示せ．

3　<u>指定された σ^2 の値にたいする χ^2 検定についての正規性の仮定からのずれ</u>　統計量

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

を考える．ここで σ_0 は σ の指定された値である．この分布は，$x_1, \cdots, x_n$ が確率変数 $X \sim N(\mu, \sigma^2)$ の観測値であるとき，$\chi^2(n-1)$ である．

3.1　X が，分散 σ^2 で $\beta_2 - 3 = \gamma_2 \neq 0$（非正規）の分布に従うとする．次式を証明せよ．

$$E(s^2) = \sigma^2, \qquad V(s^2) = \sigma^4\left(\frac{2}{n-1} + \frac{\gamma_2}{n}\right).$$

3.2

$$z = \frac{\chi^2-(n-1)}{\sqrt{2(n-1)}} = \sqrt{\frac{n-1}{2}}\left(\frac{s^2}{\sigma_0^2}-1\right)$$

の漸近分布は $\gamma_2 = 0$ のとき $N(0,1)$ で，一般には $N(0, 1+\gamma_2/2)$ である．ただし σ_0^2 は σ^2 の真の値である．

3.3　帰無仮説，$\sigma^2 = \sigma_0^2$ を検定するための $(n-1)s^2/\sigma_0^2$ に基づく χ^2 検定は，正規性からのずれにたいして頑健でないことを導け．

4　$x_i \sim G(\alpha_i, p_i)$，$i=1, \cdots, k$ で独立とする．次の仮説にたいする適当な検定を導け．

（a）　$p_1 = \cdots = p_k$ を与えたときの $\alpha_1 = \cdots = \alpha_k$

（b）　$\alpha_1 = \cdots = \alpha_k$ を与えたときの $p_1 = \cdots = p_k$

（c）　$\alpha_1 = \cdots = \alpha_k$ と $p_1 = \cdots = p_k$

[ヒント：尤度比基準を試みよ．]

5　$x_1, \cdots, x_m$ を確率密度が $\theta_1^{-1}\exp(-x/\theta_1)$ である確率変数 X の m 個の独立な観測値とし，$y_1, \cdots, y_n$ を確率密度が $\theta_2^{-1}\exp(-y/\theta_2)$ である確率変数 Y の n 個の独立な観測値とする．

5.1　$H_0: \theta_1 = \theta_2$ にたいする尤度比検定は $\bar{x}/\bar{y}$ の関数であることを示せ．$\bar{x}/\bar{y}$ の分布を求めよ．

5.2　相似検定のクラスを特徴づけ，(a) 対立仮説が $\theta_1 > \theta_2$ と (b) 対立仮説が両側（局所不偏で最強力検定）のとき最適相似検定を導け．またこれらの検定は統計量 $\bar{x}/\bar{y}$ に依存していることを示せ．[$\theta_1 = \theta_2 = \theta$ のとき $T = \sum x_i + \sum y_i$ は θ にたいして十分であることに注意し，7a.5の方法を適用せよ．]

5.3　これらの検定の検出力関数を導け．

6　$x_1, x_2, \cdots, x_m$ をポアソン分布 $p(\mu)$ からの独立な観測値とし，$f_0, f_1, \cdots$ を $0, 1, \cdots$ の度数とする．

6.1　μ にたいして十分である $T = x_1 + x_2 + \cdots + x_m$ を与えたとき，$f_0, f_1, \cdots$ の条件つき確率を求めよ．

6.2　特に T を与えたとき，観測値が 0 の度数 f_0，の正確な条件つき確率を導け．次式を証明せよ．

$$E(f_0|T) = m\left(1 - \frac{1}{m}\right)^T$$

$$V(f_0|T) = m(m-1)\left(1 - \frac{2}{m}\right)^T + m\left(1 - \frac{1}{m}\right)^T - m^2\left(1 - \frac{1}{m}\right)^{2T}.$$

6.3

$$[f_0 - E(f_0|T)] \div \sqrt{V(f_0|T)}$$

の漸近分布は $m,\ T \to \infty$ のとき $N(0,1)$ であることを示せ．

6.4　観測値が 0 の度数がポアソン分布からの抽出で期待されるものであるという仮説にたいする相似検定は，T を与えたときの f_0 の条件つき分布に基づいていることを示せ．(Rao and Chakravarthy (1956d))

6.5　$z_1, \cdots, z_{m-1}$ は，連続な分布関数 $F(x)$ の理論的な分位点，つまり $F(z_i) = i/m$，$i = 1, \cdots, m-1$ となるような値とする．$x_1, \cdots, x_T$ を $F(x)$ からの T 個の独立な標本とし，f_r を観測値 $x_1, \cdots, x_T$ のうち r 個を含む区間 $(-\infty, z_1), (z_1, z_2), \cdots, (z_{m-1}, \infty)$ の数とする．$f_0, f_1, \cdots$ の分布は問題 6.1 で導いたものと同一であり，また f_0 の分布は問題 6.2 で導いたものと同一であることを示せ．

7　<u>ノンパラメトリック検定</u>

7.1　$x_1, \cdots, x_m$ と $y_1, \cdots, y_n$ を連続分布関数 F_1, F_2 をもつ母集団からの互いに独立な観測値とする．両方を併合した標本の k 番目の順序統計量を越える x と y の個数を m_1 と n_1 で定義する．$F_1 = F_2$ ならば，与えられた m_1, n_1 の値の確率は

$$P(m_1, n_1) = \binom{m}{m_1}\binom{n}{n_1} \div \binom{m+n}{k}, \qquad m_1 + n_1 = m + n - k$$

であることを証明せよ．(m_1, n_1) のこの確率分布に基づいて，$F_1 = F_2$ なる仮説のノンパラメトリック検定を導け．

7.2　2つの皮膚軟膏，A と B の差を検定するための実験で，処理を各患者の右手と左手にランダムに施し，その効果を，悪い，普通，良いというように点をつける．50人の患者のデータは次のようにまとめられる．

		処理 A 悪い	普通	よい
処理	悪い	2	7	10
理	普通	4	4	16
B	よい	3	4	0

処理の差を検定せよ．

7.3　$x_1, \cdots, x_m$ と $y_1, \cdots, y_n$ を連続分布関数 F_1 と F_2 からの標本とする．$F_1 = F_2$ を検定するためのウィルコクソン統計量は $\sum_1^n s_i$ である．ただし，$s_1, \cdots, s_n$ は2つの標本を併合したときの $y_1, \cdots, y_n$ の順位である．$U = \sum s_i - n(n+1)/2$ として，次式を証明せよ．

$$E\left(\frac{U}{mn}\right) = \int F_1 dF_2$$

$$mnV\left(\frac{U}{mn}\right) = \int F_1 dF_2 + (n-1)\int (1-F_2)^2 dF_1 + (m-1)\int F_1^2 dF_2 - (m+n-1)\left(\int F_1 dF_2\right)^2.$$

特に，$F_1 = F_2$ のとき

$$E\left(\frac{U}{mn}\right) = \frac{1}{2}, \qquad V\left(\frac{U}{mn}\right) = \frac{m+n+1}{12mn}.$$

8　対立仮説 $\pi_1 > \pi_0$ にたいして 2 項確率 π_0 を検定するための逐次検定は，次の形に書けることを示せ．第 m 段階で m_1 を成功の数とする．このとき

(a)　$m_1 \leqq cm + d_1$ ならば π_0 を採択する．

(b)　$m_1 \geqq cm + d_2$ ならば π_0 を棄却する．

(c)　それ以外では抜取りをつづける．

ただし c, d_1, d_2 は検定の強さ (α, β) と π_0, π_1 に依存する定数である．

このS.P.R.T.のO.C.関数とA.S.N.を求めよ．

9　<u>S.P.R.T.の打ち切り</u>　S.P.R.T.が $n = 1, \cdots, N-1$ で終わらないならば，第 N 段階で検定を終了するために次の方式を採用するとしよう；$\log B < z_1 + \cdots + z_N \leqq 0$ ならば H_0 を採択し，$0 < z_1 + \cdots + z_N < \log A$ ならば H_1 を採択する．

$$y_0 = -\sqrt{N}E(z|\theta_0)/\sigma(z|\theta_0), \quad y_0' = \sqrt{N}[(1/N)\log A - E(z|\theta_0)]/\sigma(z/\theta_0)$$

とする．ただし $\sigma^2(z|\theta_0)$ は z の分散である．第1種の誤りの上限は

$$\alpha + \Phi(y_0') - \Phi(y_0) + O\left(\frac{1}{\sqrt{N}}\right)$$

であることを証明せよ．ただし Φ は $N(0,1)$ の分布関数である．β_N の上限を求めよ．

10　<u>管理図のO.C.</u>　中心線が L_0 で，L_0 の上に2つの直線 L_1, L_2 があり，L_0 の下に2つの直線 L_2', L_1' がある管理図で管理するときに，処置をとるために提案されたいくつかの方式は，次々にプロットされた点の形状に依存する次の諸方式の1つまたはその組合わせを採用することである．

(a)　点が L_1 の上または L_1' の下に落ちるとき．

(b)　2つの相連続する点が L_1 と L_2 の間，または L_1' と L_2' の間に落ちるとき．

(c)　3つの点の1番目と3番目が L_1 と L_2 の間にあり，2番目の点が L_0 と L_2 の間にあるような形状と，L_0, L_2', L_1' に関する同様な状態のとき．

$\alpha_1, \alpha_2, \alpha_3$ をプロットされた点が L_1 の上，L_2 と L_1 の間，L_0 と L_2 の間に落ちる確率とする．同様に，$\alpha_1', \alpha_2', \alpha_3'$ を L_1', L_2', L_0 に関して定義する．P_n を第 n 段階で処置をとる確率，$Q_n(L_i, L_j)$ を点が L_i と L_j の間に落ち，何の処置もとらない確率とする．

10.1　方式 (a) と (b) を同時に用いたとき，P_n, Q_n は次の漸化式を満足することを証明せよ．

$$P_n = \alpha_1 + \alpha_1' + \alpha_2 + \alpha_2' - Q_n(L_1, L_2) - Q_n(L_1', L_2')$$

$$Q_n(L_1, L_2) = \alpha_2 - \alpha_2 Q_{n-1}(L_1, L_2)$$

$$Q_n(L_1', L_2') = \alpha_2' - \alpha_2' Q_{n-1}(L_1', L_2')$$

10.2 $n \to \infty$ のとき $\lambda = \lim \sum_{r=1}^{n} P_r/n$ で，同じく $n \to \infty$ のとき，

$$p_{ij} = \lim \sum_{r=1}^{n} Q_r(L_i, L_j)/n, \quad p'_{ij} = \lim \sum_{r=1}^{n} Q_r(L'_i, L'_j)/n$$

とするとき，次式を証明せよ．

$$\lambda = \alpha_1 + \alpha'_1 + \alpha_2 + \alpha'_2 - p_{12} - p'_{12}$$

$$p_{12} = \alpha_2 - \alpha_2 p_{12}, \qquad p'_{12} = \alpha'_2 - \alpha'_2 p'_{12}.$$

λは，α_i, α'_i の値を規定する任意の仮説について，長的間に処置をとる機会の割合であることを確かめよ．

10.3 方式 (a) と (b) を同時に用いたときの O. C. 関数は

$$\lambda = \alpha_1 + \alpha'_1 + \alpha_2 + \alpha'_2 - \frac{\alpha_2}{1+\alpha_2} - \frac{\alpha'_2}{1+\alpha'_2}$$

であることを導け．

10.4 同様に方式 (a) と (c) を同時に用いたときの O. C. 関数，

$$\lambda = \alpha_1 + \alpha'_1 + \alpha_2 + \alpha'_2 - \frac{\alpha_2}{1+\alpha_2+\alpha_2\alpha_3} - \frac{\alpha'_2}{1+\alpha'_2+\alpha'_2\alpha'_3}$$

を求めよ．

10.5 各観測値を $N(\theta, \sigma^2)$ から得，$\theta = 50$ に水平線 L_0 があり，L_0 から $\pm 2\sigma$ の距離に L_2, L'_2 があり，$\pm 3\sigma$ の距離に L_1, L'_1 があるような管理図を仮定する．次の θ の値

$$50,\ 50+\frac{\sigma}{2},\ 50-\frac{\sigma}{2},\ 50+\sigma,\ 50-\sigma,\ 50+2\sigma,\ 50-2\sigma$$

にたいする問題 **10.3** と **10.4** の O. C. 関数 λ を表にせよ．

10.6 問題 **10.5** の場合，問題 **10.3** と **10.4** の方式にたいして，$\theta = 50$ での λ の値を，与えられた値（第1種の誤り）に固定するとき，直線 L_1, L_2, L'_1, L'_2 の最適配置を求めよ．読者は L_1, L'_1 や L_2, L'_2 などを対称に配置するなどのいくつかの制約を設けて，O. C. 関数をある意味で（局所的にまたは θ の指定された値で）最大にしなければならないであろう．

11 出生順に正常 (N) と欠陥 (D) というように記録された子供の系列 $NNNNDDDND$ を考えよう．欠陥児の出生順位の和は，出生の順序（出生順位）と欠陥との関連を検定するのによい基準であることを確かめよ．この問題では，和は $5+6+7+9=27$ である． 5個のNと4個のDのすべての系列が等しく起こりうるという仮説の下で，和が27以上となる確率を求めよ [Haldane and Smith (1948)].

12 得られる観測値の数に上限を設けて，H_1 にたいして仮説 H_0 の逐次検定を考えよ．[観測値の数に上限をおいて帰無仮説を検定することについては Rao (1950e) を見よ．]

13 統計的逆説 (Lindley (1957))*) $\bar{x}$ を $N(\theta, 1)$ からの n 個の観測値の平均値とし，θ の事前確率は1点 $\theta = \theta_0$ で c，θ_0 を除く θ_0 のまわりの区間では一様，であるとする．仮説 $\theta = \theta_0$ を検定したいとして，$\bar{x} = \theta_0 + \lambda_\alpha/\sqrt{n}$ が観測されるとしよう．ここで λ_α は $N(0,1)$ の α% 有意点である．$\bar{x} = \theta_0 + \lambda_\alpha/\sqrt{n}$ において $\theta = \theta_0$ である事後確率は

$$\bar{c} = ce^{-(1/2)\lambda_\alpha^2} \div [ce^{-(1/2)\lambda_\alpha^2} + (1-c)\sqrt{2\pi/n}]$$

で，$n \to \infty$ のとき $\bar{c} \to 1$ であることを示せ．したがって α を有意水準とすると，$\bar{c} = 1 - \alpha$ となる n が存在する．$\alpha = 0.05$ とすると，$(\bar{x} - \theta_0)\sqrt{n} = \lambda_\alpha$ だから，帰無仮説が5%で棄却されるような場合があるが，帰無仮説が真であるという事後確率は95%である．

*) この Lindley の論文には誤りがある．Bartlett (1957), *Biometrika* **44** 参照．[訳注]

参 考 文 献

Anscombe, F. J. (1952), Large sample theory of sequential estimation, *Proc. Camb. Phil. Soc.* 48, 600-607.

Anscombe, F. J. (1953), Sequential estimation, *J. Roy. Statist. Soc.* **B 15**, 1-29.

Bahadur, R. R. (1960), Stochastic comparison of tests, *Ann. Math. Statist.* **31**. 276-295.

Bahadur, R. R. (1958), A note on the fundamental identity of sequential analysis, *Ann. Math. Statist.* 29, 534-543.

Benerji Saibal Kumar (1960), Approximate confidence interval for linear functions of means of k populations when the population variances are not equal, *Sankhyā* **22**, 357-358.

Barnard, G. A. (1942), A new test for 2×2 tables, *Nature* **156**, 177.

Barnard, G. A. (1949), Statistical inference, *J. Roy. Statist. Soc.* **B 11**, 115-149.

Barnard, G. A. (1969), Practical applications of tests with power one, *Bull. Inst. Inter. Statist.* XLIII (i), 389-393.

Barnard, G. A., G. M. Jenkins and C. B. Winsten (1962), Likelihood inference and time series (with discussion), *J. Roy. Statist. Soc.* **A 125**, 321-372.

Basu, D. (1964), Recovery of ancillary information, *Sankhyā* **26**, 3-16.

Birnbaum, A. (1962), On the foundations of statistical inference (with discussion), *J. Am. Stat. Assn.* **57**, 269-326.

Blackwell, D. and M. A. Girshick (1954), *Theory of Games and Statistical Decisions*, Wiley, New York.

Box, G. E. P. and S. L. Anderson (1955), Permutation theory in the derivation of robust criteria and study of departures from assumptions, *J. Roy. Statist. Soc.* **B 17**, 1-34.

Chernoff, H. (1949), Asymptotic studentization in testing of hypothesis, *Ann. Math. Statist.* **20**, 268-278.

Chernoff, H. (1952), A measure of asymptotic efficiency of tests of a hypothesis based on the sum of observations, *Ann. Math. Statist.* **23**, 493-507.

Chernoff, H. (1959), Sequential design of experiments, *Ann. Math. Statist.* 30, 755-770.

Chernoff, H. and L. E. Moses (1959), *Elementary Decision Theory*, Chapman and Hall, London.

Chernoff, H. and R. I. Savage (1958), Asymptotic normality and efficiency of certain nonparametric test statistics, *Ann. Math. Statist.* **29**, 972-994.

Cochran, W. G. and G. M. Cox (1957), *Experimental Designs* (second edition), Wiley, New York.

Cox, D. R. (1958), Some problems connected with statistical inference, *Ann. Math. Statist.* **29**, 357-372.

Dantzig, G. B. (1940), On the non-existence of tests of 'Student's' hypothesis having power functions independent of σ, *Ann. Math. Statist.* **11**, 186-192.

Dantzig, G. B. and A. Wald (1951), On the fundamental lemma of Neyman and Pearson, *Ann. Math. Statist.* **22**, 87-93.

Dodge, H. F. and H. G. Romig (1929), A method of sampling inspection, *Bell System Tech. Jour.* **8**, 613-631.

Fisher, R. A. (1935a), *The Design of Experiments*, Oliver and Boyd, Edinburgh.

Fisher, R. A. (1935b), The fiducial argument in statistical inference, *Ann. Eugen.* **6**, 391-398.

Fisher, R. A. (1936), Uncertain inference, *Proc. Am. Acad. of Arts and Sciences* **71**, 245-258.

Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd, London.

Fraser, D. A. S. (1957), *Nonparametric Methods in Statistics*, Wiley, New York.

Godambe, V. P. and D. A. Sprott (1971), *Foundations of Statistical Inference*, Holt, Reinhart and Winston, Canada.

Haldane, J. B. S. and C. A. B. Smith (1948), A simple exact test for birth-order effect, *Ann. Eugen.* **14**, 117-124.

Hodges, J. L. Jr. and E. L. Lehmann (1956), The efficiency of some nonparametric competitors of the *t*-test, *Ann. Math. Statist.* **27**, 324-335.

Hoeffding, Wassily (1951), "Optimum" nonparametric tests, *Proc. (Second) Berkeley Symp. on Math. Statist. Prob.* Berkeley Univ. California Press, 83-92.

Hoeffding, Wassily and J. R. Rosenblatt (1955), The efficiency of tests, *Ann. Math. Statist.* **26**, 52-63.

Hogben, L. (1957), *Statistical Theory: the relationship of probability, credibility, and error; an examination of the contemporary crisis in statistial theory from a behaviouristic viewpoint.* George Allen and Unwin, London.

Jeffreys, H. (1948), *Theory of Probability* (second edition), Clarendon Press, Oxford.

Kempthorne, O. (1952), *Design and Analysis of Experiments*, Wiley, New York.

Kyburg, H. E., Jr. (1961), *Probability and the Logic of Rational Belief*, Wesleyan Univ. Press. Middletown, Connecticut.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, Wiley, New York.

Lindley, D. V. (1953), Statistical inference, *J. Roy. Stat. Soc.* **B 15**, 30-76.

Lindley, D. V. (1957), A statistical paradox, *Biometrika* **44**, 187-192.

Linnik, Yu. V. (1963), On the Behrens-Fisher problem, *Proc. Int. Stat. Conference*, Ottawa.

Mahalanobis, P. C. (1940), A sample survey of acreage under jute in Bengal with discussion on planning of experiments, *Proc. (Second) Indian Stat. Conference, Calcutta.*

Mitra, Sujit Kumar (1960), On the F-test in the intrablock analysis of a balanced incomplete block design, *Sankhyā* **17**, 279-284.

Neyman, J. (1935), On the problem of confidence intervals, *Ann. Math. Statist.* **6**, 111-116.

Neyman, J. (1937), Outline of a theory of statistical estimation based on the classical theory of probability, *Philos. Trans. Roy. Soc.* **A 236**, 333-380.

Neyman, J. (1961), Silver jubilee of my dispute with Fisher, *J. Operations Research Soc. (Japan)* **3**, 145-154.

Neyman, J. with the cooperation of K. Iwaszkiewicz and St. Kolodzieszyk (1935), Statistical problems in agricultural experimention, *J. Roy. Statist. Soc.* (suppl.) **2**, 107-154.

Neyman, J. and E. S. Pearson (1933), On the problem of the most efficient tests of statistical hypotheses, *Phil. Trans. Roy. Soc.* **A 231**, 289-337.

Noether, G. E. (1955), On a theorem of Pitman, *Ann. Math. Statist.* **26**, 64-68.

Pitman, E. J. G. (1937), Significance tests which may be applied to samples from any population, *J. Roy. Statist. Soc. (suppl.)* **4**, 119-130, 225-232, and *Biometrika* **29**, 322-335.

Pitman E. J. G. (1949), Lecture notes on nonparametric statistical inference, Columbia University.

Plackett, R. L. (1960), *Principles of Regression Analysis*, Clarendon Press, Oxford.

Puri, M. L. and P. K. Sen (1970), *Nonparametric Methods in Multivariate Analysis*, Wiley, New York.

Robbins, H. and D. Siegmund (1969), Confidence sequences and interminable ests, *Bull. Inst. Int. Statist.* XLIII (1), 379-387.

Savage, I. R. (1953), Bibliography of nonparametric statistics and related topics, *J. Am. Stat. Assoc.* **48**, 844-906.

Savage, L. J. (1954), *Foundation of Statistics*, Wiley, New York.

Savage, L. J. (1962), *Foundations of Statistical Inference*, A discussion, Methuen, London.

Scheffe, H. (1959), *The Analysis of Variance*, Wiley, New York.

Siegel, Sidney (1956), *Nonparametric Statistics for the Behavioural Sciences*, McGraw-Hill, New York, Toronto, London.

Stein, Charles (1945), A two-samples test for a linear hypothesis whose power is independent of the variance, *Ann. Math. Satist.* **16**, 243-258.

Van der Waerden, B. L. and E. Nievergelt (1956), *Tables for Comparing Two Samples by X-Test and Sign Test*, Springer-Verlag, Berlin.

Wald, A. (1941), Asymptotically most powerful tests of statistical hypotheses, *Ann. Math. Statist.* **12**, 1-19.

Wald, A. (1947), *Sequential Analysis*, Wiley, New York.

Wald, A. (1950), *Statistical Decision Functions*, Wiley, New York.

Wald, A. and J. Wolfowitz (1940), On a test whether two samples are from the same population, *Ann. Math. Statist.* **11**, 147-162.

Wash, J. E. (1962), *Handbook of Nonparametric Statistics. Investigation of Randomness, Moments, Percentiles and Distributions*, Nostrand, London.

Welch, W. L. (1947), The generalization of Student's problem when several population variances are involved, *Biometrika* **34**, 28-35.

Wilcoxon, Frank (1945), Individual comparisons by ranking methods, *Biometrics* **1**, 80-83.

Wolfowitz, J. (1947), The efficiency of sequential estimates and Wald's equation for sequential processes, *Ann. Math. Statist.* **18**, 215-230.

Wolfowitz, J. (1949), Nonparametric statistical inference, *Proc. (First) Berkeley Symp. on Math. Satist. Prob.* Berkeley Univ. California. 93-113.

第8章　多変量解析

まえがき　多変量分布と条件つき分布の一般的概念は，第2章で述べた．1つの確率変数の，他の確率変数を与えたときの，条件つき分布の平均値であるところの回帰の性質については，**4g.1**から**4g.2**で学んだ．この章では多変量正規分布として知られている特別な種類の分布について考えるが，この分布は多変量測定値が得られたときの統計的推測において重要な役割を演ずるものである．

多変量正規分布は，互いに独立な1変量正規確率変数のいくつかの1次関数の分布として，すでに**3b.3**で導入した．p変量正規分布ですべての分散が等しくかつすべての共分散が等しいという特殊な場合と，一般の2変量正規分布については，それぞれ**3c.1**と**3d.1**において詳細に考察した．この章では，一般のp変量正規分布の性質を調べ，いくつかの標本統計量の分布を導くとともに，それらの実際問題における使い方を例示する．

この章におけるp変量正規分布への理論的接近は，通常の方法には従わないので，若干の注意が必要である．第1に，その分布を確率密度関数によっては定義しないで，p変数のどの1次関数も1変量正規分布に従うという性質によって特徴づける．第2に，このような特徴づけを標本統計量の分布を導くのに利用する．1変量理論における既知の結果に対応して，多変量理論への一般化を，さらに進んだ解析をほとんどすることなく書き下せることを示す．これを行う方法は，**8b.1**と**8b.2**で述べる．たとえば，1変量理論における標本平均と標本分散の同時分布を知ることによって，多変量測定値があるときの標本平均と標本分散・共分散の同時分布を書き下すことができる．分散・共分散分析による多変量有意性検定の理論の全体が，1変量分散分析の一般化として得られる．読者はこの一般的接近法に習熟すべきである．

多変量正規分布について他にも多くの特徴づけがなされているが，それらは後の結果の導出に用いられないので，初めて読むときは省いてもよいだろう．しかしこれらの特徴づけはヒルベルト空間やより一般的な空間における正規分布の理論を勉強したい人には役立つであろう．

統計学の文献ではあまり強調されていないけれども，多変量解析の重要な側面は，いくつかの特性の測定値がすでに得られているときに，それに2,3の特性のデータを加えることによって得られる情報の量を吟味することである．これは重要なことである．どんな問

題においても，測定できる特性の数は実際にはかぎりがないので，あるいくつかの特性値が存在するとき，他の特性値が余分であるかどうかを検討することは必要である．この問題については，8c.4 で検討する．多くの実際問題は 8d.1 から 8d.5 で考察する．

8a　多変量正規分布

8a.1　定　　義

多変量確率密度については，3b.3 において，独立な 1 変量正規確率変数の線形関数の標本分布を学ぶ際にすでに述べた．p 次元確率変数 $(U_1, \cdots, U_p)$ を $\mathbf{U}$ と書き，（便宜上）同じ記号で，その変数のとり得る値をも示すとすると，p 変量正規分布の密度関数は

$$(2\pi)^{-p/2}|\mathbf{A}|^{1/2}\exp\left[-\frac{1}{2}(\mathbf{U}-\boldsymbol{\mu})'\mathbf{A}(\mathbf{U}-\boldsymbol{\mu})\right] \tag{8a.1.1}$$

で定義された．ここで $\mathbf{A}$ は正定符号行列である．また $E(\mathbf{U})=\boldsymbol{\mu}$ および $D(\mathbf{U})=\mathbf{A}^{-1}$ であることがわかる．ここで，$D(\mathbf{U})$ は $\mathbf{U}$ の成分の**分散共分散行列**である．分散共分散行列を dispersion 行列とよび，分散共分散を導く演算子として記号 D を用いる．

しかし，本書では密度関数 (8a.1.1) を多変量正規分布の研究の基礎に選ばない．そのかわりに，可算または非可算の次元をもつより複雑な確率変数へその概念を拡張できるような方法で分布を特徴づける．さらに，(8a.1.1) のような密度関数は $D(\mathbf{U})$ が特異ならば存在しないが，一般的な接近法ではそのような場合も含める必要がある．

この目的のために，p 次元確率変数の分布は，$\mathbf{T}$ を任意の実定数ベクトルとすると，線形関数 $\mathbf{T}'\mathbf{U}$ の 1 次元分布によって完全に決定されるというクラメールとウォルドの結果を利用しよう．その結果を証明するために，$\phi(t, \mathbf{T})$ を $\mathbf{T}'\mathbf{U}$ の特性関数とすると，

$$\phi(t, \mathbf{T}) = E[\exp(it\mathbf{T}'\mathbf{U})] \tag{8a.1.2}$$

から，$\phi(1, \mathbf{T}) = E[\exp(i\mathbf{T}'\mathbf{U})]$ が得られ，これは $\mathbf{T}$ の関数と考えると，$\mathbf{U}$ の特性関数にほかならない．それゆえ，$\mathbf{U}$ の分布は特性関数の反転定理によって，一意に定義される．この結果から，p 変量正規分布を次のように定義する．

定義 1.　p 次元確率変数 $\mathbf{U}$，すなわち，E_p（p 次元ユークリッド空間）において値をとる確率変数 $\mathbf{U}$ は，$\mathbf{U}$ のすべての線形関数が 1 変量正規分布に従うならば，そしてそのときにのみ，p 変量正規分布 N_p に従うという．

(8a.1.2) の結果は，定義 1 を満足する確率変数 $\mathbf{U}$ が存在するならば，その分布は一意的に決定されることを示すに過ぎない．しかし，8a.2 で，この定義に従う N_p が存在す

ることを証明しよう．

N_p に関する定義 1 は，すべての線形汎関数の標本分布が 1 変量正規分布であることを要求することによって，ヒルベルトまたはバナッハ空間のような，より一般的な空間における正規確率測度を定義することへ拡張できることがわかる (Fréchet (1951))．この章では，有限次元確率変数のみに注目する．しかし，議論する問題の取り扱いと性質は，応用研究の多くの分野で考えられる，より一般的な確率変数の分布の研究に興味をもつ人にとってもよい準備として役立つであろう．(Grenander and Rosenblatt (1957), Prohorov and Fisz(1957), Rao and Varadarajan (1963b), Wold (1938) などを見よ．)

記号と演算　ベクトル確率変数についての以下の記号と演算は，種々の命題の証明に用いられる．$\mathbf{X}' = (X_1, \cdots, X_p)$ および $\mathbf{Y}' = (Y_1, \cdots, Y_q)$ を 2 つの確率変数とする．E, D, C, V によって，それぞれ期待値，分散共分散行列，共分散および分散をあらわすと，

$$E(\mathbf{X}') = [E(X_1), \cdots, E(X_p)]$$

$$D(\mathbf{X}) = \begin{pmatrix} V(X_1) & \cdots & C(X_1, X_p) \\ & \cdots\cdots\cdots\cdots & \\ C(X_p, X_1) & \cdots & V(X_p) \end{pmatrix} \tag{8a.1.3}$$

$$C(\mathbf{X}, \mathbf{Y}) = \mathrm{cov}(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} C(X_1, Y_1) & \cdots & C(X_1, Y_q) \\ & \cdots\cdots\cdots & \\ C(X_p, Y_1) & \cdots & C(X_p, Y_q) \end{pmatrix} \tag{8a.1.4}$$

$$C(\mathbf{Y}, \mathbf{X}) = [C(\mathbf{X}, \mathbf{Y})]'$$

$$V(\mathbf{L}'\mathbf{X}) = \mathbf{L}'D(\mathbf{X})\mathbf{L}, \qquad C(\mathbf{L}'\mathbf{X}, \mathbf{M}'\mathbf{Y}) = \mathbf{L}'C(\mathbf{X}, \mathbf{Y})\mathbf{M} \tag{8a.1.5}$$

となる．ここで $\mathbf{L}$ および $\mathbf{M}$ は列ベクトルである．

$\mathbf{X}_1, \cdots, \mathbf{X}_k$ を k 個の p 次元確率変数とし，$A_1, \cdots, A_k$ を固定した定数とすれば，

$$\begin{aligned} D(A_1\mathbf{X}_1 + \cdots + A_k\mathbf{X}_k) &= \sum A_i^2 D(\mathbf{X}_i) + 2\sum\sum A_iA_jC(\mathbf{X}_i, \mathbf{X}_j) \\ &= \sum A_i^2 D(\mathbf{X}_i), \quad \mathbf{X}_i \text{ が無相関のとき} \\ &= (\textstyle\sum A_i^2)D(\mathbf{X}), \quad \mathbf{X}_i \text{ が無相関で等分散共分散をもつとき} \end{aligned} \tag{8a.1.6}$$

となる．$\mathbf{X}$ を p 次元確率変数，$\mathbf{B}$ を $q \times p$ 行列とすれば，$\mathbf{BX}$ は q 次元確率変数となり，

$$\left.\begin{aligned} E(\mathbf{BX}) &= \mathbf{B}E(\mathbf{X}) \\ D(\mathbf{BX}) &= \mathbf{B}D(\mathbf{X})\mathbf{B}' \end{aligned}\right\} \tag{8a.1.7}$$

を得る．行列 $\mathbf{A}$ の階数を rank $\mathbf{A}$ または $R(\mathbf{A})$ であらわす．

8a.2　分 布 の 性 質

定義 1 の意味で $\mathbf{U} \sim N_p$ とすると，以下の結果が成立する．

(i)　$E(\mathbf{U})$ および $D(\mathbf{U})$ が存在する．それぞれを $\boldsymbol{\mu}$ および $\boldsymbol{\Sigma}$ であらわすと，定数ベクトル $\mathbf{T}$ にたいして，$\mathbf{T}'\mathbf{U}$ は平均 $\mathbf{T}'\boldsymbol{\mu}$ と分散 $\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}$ をもつ 1 変量正規確率変数となる．

すなわち，$\mathbf{T}'\mathbf{U} \sim N_1(\mathbf{T}'\boldsymbol{\mu}, \mathbf{T}'\boldsymbol{\Sigma}\mathbf{T})$ となる．

$\mathbf{U}$ の成分を $(U_1, \cdots, U_p)$ とする．定義によって，U_i は1変量正規であるので，$E(U_i) < \infty$ かつ $V(U_i) < \infty$ である．$V(U_i)$ と $V(U_j)$ が存在するので，$\mathrm{Cov}(U_i, U_j)$ も存在する．$E(U_i) = \mu_i$，$V(U_i) = \sigma_{ii}$，$\mathrm{cov}(U_i, U_j) = \sigma_{ij}$ とすると，

$$E(\mathbf{T}'\mathbf{U}) = \mathbf{T}'\boldsymbol{\mu}, \quad \text{ここで } \boldsymbol{\mu}' = (\mu_1, \cdots, \mu_p)$$
$$V(\mathbf{T}'\mathbf{U}) = \mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}, \quad \text{ここで } \boldsymbol{\Sigma} = (\sigma_{ij}), \tag{8a.2.1}$$

したがって $\mathbf{T}'\mathbf{U} \sim N_1(\mathbf{T}'\boldsymbol{\mu}, \mathbf{T}'\boldsymbol{\Sigma}\mathbf{T})$．

（ii） $\mathbf{U}$ の特性関数は次のようになる．

$$\exp\left(i\,\mathbf{T}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}\right). \tag{8a.2.2}$$

1変量正規確率変数 $u \sim N_1(\mu, \sigma^2)$ の特性関数は（95 ページ参照）

$$\exp\left(it\mu - \frac{1}{2}t^2\sigma^2\right) \tag{8a.2.3}$$

であり，(8a.2.3) を用いて，確率変数 $\mathbf{T}'\mathbf{U} \sim N_1(\mathbf{T}'\boldsymbol{\mu}, \mathbf{T}'\boldsymbol{\Sigma}\mathbf{T})$ の特性関数は

$$\exp\left(it\mathbf{T}'\boldsymbol{\mu} - \frac{1}{2}t^2\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}\right) \tag{8a.2.4}$$

となる．(8a.2.4) において $t = 1$ とおけば，$\mathbf{U}$ の特性関数は (8a.2.2) となることがわかる．以下は (8a.2.2) の直接の結果である．

p 変量正規分布は，特性関数 (8a.2.2) が $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ のみを含むので，確率変数の平均ベクトル $\boldsymbol{\mu}$ と分散共分散行列 $\boldsymbol{\Sigma}$ によって完全に定めることができる．したがって，p 変量正規分布は $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ をパラメータとして $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ と書ける．

（a） 任意の $\mathbf{T}$ について，$\mathbf{T}'\mathbf{U} \sim N_1(\mathbf{T}'\boldsymbol{\mu}, \mathbf{T}'\boldsymbol{\Sigma}\mathbf{T})$ となるベクトル $\boldsymbol{\mu}$ と行列 $\boldsymbol{\Sigma}$ が存在すれば，$\mathbf{U} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ である．

（b） $\boldsymbol{\Sigma} = \boldsymbol{\Delta}$（対角行列）であれば，成分 $U_1, \cdots, U_p$ は独立で，それぞれ1変量正規である．

このとき特性関数 (8a.2.2) は p 個の要素の積となり，変数の独立性を意味することに注意せよ．すなわち，すべての積率相関の値が 0 であることは $\mathbf{U}$ の成分が独立であることの必要十分条件である．

（c） $\mathbf{U}_1$ と $\mathbf{U}_2$ が変量 $\mathbf{U}$ の2つの部分集合であるとすると，

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} D(\mathbf{U}_1) & C(\mathbf{U}_1, \mathbf{U}_2) \\ C(\mathbf{U}_2, \mathbf{U}_1) & D(\mathbf{U}_2) \end{pmatrix} \tag{8a.2.5}$$

と書くことができる．ここで，$\boldsymbol{\Sigma}_{11}$ と $\boldsymbol{\Sigma}_{22}$ は $\mathbf{U}_1, \mathbf{U}_2$ の分散共分散行列であり，$\boldsymbol{\Sigma}_{21}$ は $\mathbf{U}_2, \mathbf{U}_1$ の共分散行列（すなわち，$\mathbf{U}_1$ の成分と $\mathbf{U}_2$ の成分の共分散の行列）である．確率変数 $\mathbf{U}_1, \mathbf{U}_2$ は $\boldsymbol{\Sigma}_{12} = 0$ のとき，およびそのときに限り独立に分布する．

この結果は，(8a.2.2) のなかの2次形式を，$\mathbf{T}$ の成分から成る部分集合 $\mathbf{T}_1, \mathbf{T}_2$ を用いて，

$$\begin{aligned} \mathbf{T}'\boldsymbol{\Sigma}\mathbf{T} &= \mathbf{T}_1'\boldsymbol{\Sigma}_{11}\mathbf{T}_1 + 2\mathbf{T}_1'\boldsymbol{\Sigma}_{12}\mathbf{T}_2 + \mathbf{T}_2'\boldsymbol{\Sigma}_{22}\mathbf{T}_2 \\ &= \mathbf{T}_1'\boldsymbol{\Sigma}_{11}\mathbf{T}_1 + \mathbf{T}_2'\boldsymbol{\Sigma}_{22}\mathbf{T}_2, \quad \boldsymbol{\Sigma}_{12} = 0 \text{ のとき} \end{aligned}$$

と書くことができることから導ける．このとき，$\mathbf{U}$ の特性関数は2つの要素の積と見られる．この結果は，多変量正規確率変数の2つの部分集合の独立性は，部分集合間の共分散行列を調べることによって推論できるという点で，重要である．

（d） $\mathbf{U}$ の部分集合 $\mathbf{U}_1, \mathbf{U}_2, \cdots, \mathbf{U}_k$ は，2つずつ互いに独立であれば，それらは相互に独立である．

（e） 関数 (8a.2.2) は実際に，ある確率変数の特性関数となるので，定義1の N_p は存在する．

$\boldsymbol{\Sigma}$ は非負定符号（分散共分散行列はそうである）であるので，2次形式 $\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T}$ は

$$\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T} = (\mathbf{B}_1'\mathbf{T})^2 + \cdots + (\mathbf{B}_m'\mathbf{T})^2$$

と書くことができる（34ページ参照）．ここで $\mathbf{B}_1, \cdots, \mathbf{B}_m$ は1次独立ベクトルであり，$m = R(\boldsymbol{\Sigma})$ である．関数 (8a.2.2) は

$$e^{i\mathbf{T}'\boldsymbol{\mu}} \cdot e^{-(\mathbf{B}_1'\mathbf{T})^2/2} \cdot \cdots \cdot e^{-(\mathbf{B}_m'\mathbf{T})^2/2} \tag{8a.2.6}$$

と書くことができて，もし (8a.2.6) の各要素が特性関数であれば，特性関数の積はまた特性関数であるので，求める結果が証明される．

(8a.2.6) の第1の要素は次の p 次元確率変数の特性関数である．

$$\mathbf{Y}_0' = (\mu_1, \cdots, \mu_p) \qquad \text{確率 1 で} \tag{8a.2.7}$$

要素 $\exp\left[-\frac{1}{2}(\mathbf{B}_i'\mathbf{T})^2\right]$ は p 次元確率変数

$$\mathbf{Y}_i' = (B_{1i}G_i, \cdots, B_{pi}G_i), \qquad G_i \sim N_1(0, 1), \tag{8a.2.8}$$

の特性関数であることは容易に示される．ここで $\mathbf{B}_i' = (B_{1i}, \cdots, B_{pi})$．この数字 m は分布 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ の階数 rank と呼ばれる．

（f） $\mathbf{U} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ でその階数が m であるための必要十分条件は

$$\mathbf{U} = \boldsymbol{\mu} + \mathbf{B}\mathbf{G}, \qquad \mathbf{B}\mathbf{B}' = \boldsymbol{\Sigma} \tag{8a.2.9}$$

となることである．ここで $\mathbf{B}$ は階数 m の $p \times m$ 行列であり，$\mathbf{G} \sim N_m(\mathbf{0}, \mathbf{I})$ である．すなわち，成分 $G_1, \cdots, G_m$ は独立で，それぞれ $N_1(0, 1)$ で分布する．

[(8a.2.9) の他の導き方については **8g** を見よ．]

十分条件を証明するために，$\mathbf{U}$ の1次関数を

$$\mathbf{T}'\mathbf{U} = \mathbf{T}'\boldsymbol{\mu} + \mathbf{T}'\mathbf{B}\mathbf{G}$$

とする．ただし，$\mathbf{G} \sim N_m(\mathbf{0}, \mathbf{I})$ であり，1変量理論から，独立な $N_1(0, 1)$ 変数の1次

関数 $(\mathbf{T}'\mathbf{B})\mathbf{G}$ は $N_1(0, \mathbf{T}'\mathbf{B}\mathbf{B}'\mathbf{T})$ に従って分布する．したがって，

$$\mathbf{T}'\mathbf{U} = \mathbf{T}'\mu + \mathbf{T}'\mathbf{B}\mathbf{G} \sim N_1(\mathbf{T}'\mu, \mathbf{T}'\mathbf{B}\mathbf{B}'\mathbf{T})$$
$$\Rightarrow \mathbf{U} \sim N_p(\mu, \mathbf{B}\mathbf{B}') = N_p(\mu, \Sigma), \qquad \mathbf{T}\text{ は任意であるので.}$$

必要条件を証明するために，(8a. 2. 6)の分解式より $\mathbf{U}$ の分布は (8a. 2. 7) と (8a. 2. 8) で定義した独立な p 次元確率変数

$$\mathbf{Y}_0, \mathbf{Y}_1, \cdots, \mathbf{Y}_m$$

の分布のたたみこみであることに注意する．そうすれば $\mathbf{U}$ は

$$\mathbf{U} = \mathbf{Y}_0 + \mathbf{Y}_1 + \cdots + \mathbf{Y}_m$$
$$= \mu + G_1\mathbf{B}_1 + \cdots + G_m\mathbf{B}_m = \mu + \mathbf{B}\mathbf{G}$$

と書くことができる．ここで $\mathbf{B}_i$ と G_j は (8a. 2. 8) で定義したものと同様であり，$\mathbf{B}$ は分割行列 $(\mathbf{B}_1 \vdots \cdots \vdots \mathbf{B}_m)$ である．こうして，多変量正規分布を別の形で定義することができる．

定義 2.　p 次元確率変数 $\mathbf{U}$ は，$\mathbf{U} = \mu + \mathbf{B}\mathbf{G}$ と表現できるならば，正規分布 N_p に従うといえる．ただし $\mathbf{B}$ は階数 m の $p \times m$ 行列であり，$\mathbf{G}$ はそれぞれ平均 0，分散 1 の互いに独立な N_1（1 変量正規）確率変数から成る $m \times 1$ ベクトルである．

注 1.　$\mathbf{U} = \mu + \mathbf{B}\mathbf{G}$ の関係は，確率 1 で，確率ベクトル $\mathbf{U} \in \mathscr{M}(\mu : \Sigma)$ すなわち $\mathbf{U}$ は μ と Σ の列によって形成される線形部分空間に属することを示している．

注 2.　$\mathbf{U}_i \sim N_{p_i}(\mu_i, \Sigma_i)$，$i = 1, \cdots, k$ は独立で，T は $\mathbf{U}_1, \cdots, \mathbf{U}_k$ の関数（統計量）とする．$\mathbf{U}_i = \mu_i + \mathbf{B}_i\mathbf{G}_i$ の表現を用いて，

$$T(\mathbf{U}_1, \cdots, \mathbf{U}_k) = T(\mu_1 + \mathbf{B}_1\mathbf{G}_1, \cdots, \mu_k + \mathbf{B}_k\mathbf{G}_k)$$
$$= t(\mathbf{G}_1, \cdots, \mathbf{G}_k)$$

となる．そうすれば $\mathbf{U}_1, \cdots, \mathbf{U}_k$ からつくられる統計量 T の研究は，独立な 1 変量正規確率変数 $\mathbf{G}_1, \cdots, \mathbf{G}_k$ の関数の研究に帰せられる．このような簡約化は 1 変量理論における既知の結果が，多変量理論における結果を導くのにそのまま適用できるので役に立つ．

注 3.　定義 2 における $\mu, \mathbf{B}, \mathbf{G}$ をもつ，$\mathbf{U} = \mu + \mathbf{B}\mathbf{G}$ と，$\mathbf{C}$ が $p \times q$ 行列で $\mathbf{F}$ が独立な $N(0,1)$ 変量から成る q 次元ベクトルである $\mathbf{U} = \mu + \mathbf{C}\mathbf{F}$ は，もし $\mathbf{B}\mathbf{B}' = \mathbf{C}\mathbf{C}'$ であれば同じ分布をすることは容易に理解できる．したがって，q や $R(\mathbf{C})$ についての制限は何もない．しかし，定義におけるように $\mathbf{B}$ についての制限つきの表現は実際に応用する際に役に立つ．

（iii）　もし $\mathbf{U} \sim N_p$ であれば，$\mathbf{U}$ の q 成分から成るいかなる部分集合の周辺分布も N_q である．［定義 1 に従えば，部分集合のどの 1 次関数もまた 1 変量正規であるので.］

（iv）　$\mathbf{U}$ の q 個の 1 次関数の同時分布は，定義 1 より N_q である．$\mathbf{C}$ を $q \times p$ の行列として，$\mathbf{Y} = \mathbf{C}\mathbf{U}$ が q 個の 1 次関数をあらわすとすれば

$$\mathbf{Y} \sim N_q(\mathbf{C}\mu, \mathbf{C}\Sigma\mathbf{C}') \tag{8a. 2. 10}$$

である．

1 次関数 $\mathbf{L}'\mathbf{Y} = \mathbf{L}'\mathbf{C}\mathbf{U} = (\mathbf{L}'\mathbf{C})\mathbf{U}$ を考える．(i) から

$$(\mathbf{L}'\mathbf{C})\mathbf{U} \sim N_1[(\mathbf{L}'\mathbf{C})\mu, (\mathbf{L}'\mathbf{C})\Sigma(\mathbf{C}'\mathbf{L})]$$

すなわち，

$$\mathbf{L}'\mathbf{Y} \sim N_1(\mathbf{L}'\mathbf{C}\mu, \mathbf{L}'\mathbf{C}\Sigma\mathbf{C}'\mathbf{L}) \Rightarrow \mathbf{Y} = N_q(\mathbf{C}\mu, \mathbf{C}\Sigma\mathbf{C}').$$

（v）　$\mathbf{U}_1' = (U_1, \cdots, U_r)$，$\mathbf{U}_2' = (U_{r+1}, \cdots, U_p)$ は $\mathbf{U}$ の 2 つの部分集合であり，$\Sigma_{11}, \Sigma_{12}, \Sigma_{22}$ は (8a. 2. 5) で定義されたような Σ の分割であるとする．$\mathbf{U}_1$ が与えられたときの $\mathbf{U}_2$ の条件つき分布は，

$$N_{p-r}(\mu_2 + \Sigma_{21}\Sigma_{11}^{-}(\mathbf{U}_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-}\Sigma_{12}) \tag{8a. 2. 11}$$

となる．ここで $E(\mathbf{U}_i) = \mu_i$，$i = 1, 2$，および Σ_{11}^{-} は Σ_{11} の一般逆行列である．

行列の列によって張られた線形空間を $\mathscr{M}$ であらわすと，$\mathscr{M}(\Sigma_{12}) \subset \mathscr{M}(\Sigma_{11})$ を読者は証明できるであろう．(Σ_{11} の列への任意の直交ベクトルは同時に Σ_{12} の列にも直交することを示すことによって証明できる.) したがって，$\Sigma_{21} = \mathbf{B}\Sigma_{11}$ となるような行列 $\mathbf{B}$ が存在する．いま，一般逆行列の性質 $\Sigma_{11}\Sigma_{11}^{-}\Sigma_{11} = \Sigma_{11}$ を使って，

$$\Sigma_{21}\Sigma_{11}^{-}\Sigma_{11} = \mathbf{B}\Sigma_{11}\Sigma_{11}^{-}\Sigma_{11} = \mathbf{B}\Sigma_{11} = \Sigma_{21} \tag{8a. 2. 12}$$

となる．Σ_{11}^{-} が真の逆行列であれば，この結果は明らかである．一般逆行列に馴染みのない読者は Σ_{11}^{-1} の存在を仮定して (8a. 2. 12) を導く議論を避けてよい．

共分散行列：

$$\begin{aligned} &C[\mathbf{U}_2 - \mu_2 - \Sigma_{21}\Sigma_{11}^{-}(\mathbf{U}_1 - \mu_1), \mathbf{U}_1 - \mu_1] \\ &\quad = C(\mathbf{U}_2 - \mu_2, \mathbf{U}_1 - \mu_1) - \Sigma_{21}\Sigma_{11}^{-}C(\mathbf{U}_1 - \mu_1, \mathbf{U}_1 - \mu_1) \\ &\quad = \Sigma_{21} - \Sigma_{21}\Sigma_{11}^{-}\Sigma_{11} = \mathbf{0} \quad \text{(8a. 2. 12) より} \end{aligned} \tag{8a. 2. 13}$$

同様に

$$\begin{aligned} &D[\mathbf{U}_2 - \mu_2 - \Sigma_{21}\Sigma_{11}^{-}(\mathbf{U}_1 - \mu_1)] \\ &\quad = C(\mathbf{U}_2 - \Sigma_{21}\Sigma_{11}^{-}\mathbf{U}_1, \mathbf{U}_2 - \Sigma_{21}\Sigma_{11}^{-}\mathbf{U}_1) \\ &\quad = C(\mathbf{U}_2, \mathbf{U}_2) - C(\mathbf{U}_2, \mathbf{U}_1)\Sigma_{11}^{-}\Sigma_{12} \\ &\qquad - \Sigma_{21}\Sigma_{11}^{-}C(\mathbf{U}_1, \mathbf{U}_2) + \Sigma_{21}\Sigma_{11}^{-}C(\mathbf{U}_1, \mathbf{U}_1)\Sigma_{11}^{-}\Sigma_{12} \\ &\quad = \Sigma_{22} - 2\Sigma_{21}\Sigma_{11}^{-}\Sigma_{12} + \Sigma_{21}\Sigma_{11}^{-}\Sigma_{11}\Sigma_{11}^{-}\Sigma_{12} \\ &\quad = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-}\Sigma_{12}, \quad \because\ \Sigma_{21}\Sigma_{11}^{-}\Sigma_{11} = \Sigma_{21}. \end{aligned} \tag{8a. 2. 14}$$

$E[\mathbf{U}_2 - \mu_2 - \Sigma_{21}\Sigma_{11}^{-}(\mathbf{U}_1 - \mu_1)] = \mathbf{0}$ であるので，(8a. 2. 14) の結果より

$$\mathbf{U}_2 - \mu_2 - \Sigma_{21}\Sigma_{11}^{-}(\mathbf{U}_1 - \mu_1) \sim N_{p-r}(\mathbf{0}, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-}\Sigma_{12}) \tag{8a. 2. 15}$$

となる．

(8a. 2. 13) の結果は，$\mathbf{U}_2 - \mu_2 - \Sigma_{21}\Sigma_{11}^{-}(\mathbf{U}_1 - \mu_1)$ と $\mathbf{U}_1 - \mu_1$ は独立に分布することを

示している．したがって，(8a. 2. 15) は $\mathbf{U}_1$ が与えられた場合の $\mathbf{U}_2-\boldsymbol{\mu}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^-(\mathbf{U}_1-\boldsymbol{\mu}_1)$ の条件付分布と解釈できる．これは $\mathbf{U}_1$ が与えられた場合の $\mathbf{U}_2$ の条件付分布 (8a. 2. 11) と同値である．(8a. 2. 13) および (8a. 2. 14) の計算は $\boldsymbol{\Sigma}_{11}$ の真の逆行列が存在するときはより単純となる．

(8a. 2. 15) の特殊な場合で重要なものの1つは，U_p と $\mathbf{U}_1'=(U_1,\cdots,U_{p-1})$ のように，$\mathbf{U}_2$ が単一の成分となる場合である．$\mathbf{U}_1$ が与えられた場合の U_p の条件付分布は，$\mu_p+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^-(\mathbf{U}_1-\boldsymbol{\mu}_1)$，すなわち

$$\alpha+\beta_1U_1+\cdots+\beta_{p-1}U_{p-1} \tag{8a. 2. 16}$$

の形の $U_1,\cdots,U_{p-1}$ の1次結合を平均 ($U_1,\cdots,U_{p-1}$ への U_p の回帰) としてもち，分散 (残差分散) が $\sigma_{pp}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^-\boldsymbol{\Sigma}_{12}$ であるような1変量正規分布である．$\boldsymbol{\Sigma}_{11}$ が正則であれば (1b. 8の終わりにある問題2. 4により)

$$|\boldsymbol{\Sigma}_{11}|(\sigma_{pp}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})=|\boldsymbol{\Sigma}|$$

であるから，残差分散は次のように陽にあらわすことができる．

$$\sigma_{pp}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}=\frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}_{11}|}=\frac{1}{\sigma^{pp}}. \tag{8a. 2. 17}$$

ただし，σ^{pp} は $\boldsymbol{\Sigma}$ の逆行列の最後の要素である．

(vi) N_p の再生性　$\mathbf{U}_i\sim N_p(\boldsymbol{\mu}_i,\Sigma_i)$，$i=1,\cdots,n$ は，すべて独立とすると，定数 $A_1,\cdots,A_n$ にたいして

$$\mathbf{Y}=A_1\mathbf{U}_1+\cdots+A_n\mathbf{U}_n\sim N_p(A_i\boldsymbol{\mu}_i,\textstyle\sum A_i^2\Sigma_i). \tag{8a. 2. 18}$$

(8a. 2. 18) の証明には，線形関数 $\mathbf{L}'\mathbf{Y}$ を考えればよい．すなわち，

$$\mathbf{L}'\mathbf{Y}=A_1(\mathbf{L}'\mathbf{U}_1)+\cdots+A_n(\mathbf{L}'\mathbf{U}_n)$$

は，$\mathbf{L}'\mathbf{U}_1,\cdots,\mathbf{L}'\mathbf{U}_n$ がすべて独立で1変量正規であるので，1変量正規である．ゆえに $\mathbf{Y}\sim N_p$．さらに

$$E(\mathbf{L}'\mathbf{Y})=A_1\mathbf{L}'\boldsymbol{\mu}_1+\cdots+A_n\mathbf{L}'\boldsymbol{\mu}_n=\mathbf{L}'(A_1\boldsymbol{\mu}_1+\cdots+A_n\boldsymbol{\mu}_n)$$
$$V(\mathbf{L}'\mathbf{Y})=A_1^2V(\mathbf{L}'\mathbf{U}_1)+\cdots+A_n^2V(\mathbf{L}'\mathbf{U}_n)$$
$$=\textstyle\sum A_i^2\mathbf{L}'\Sigma_i\mathbf{L}=\mathbf{L}'(\sum A_i^2\Sigma_i)\mathbf{L}$$

は，$\sum A_i\boldsymbol{\mu}_i$ および $\sum A_i^2\Sigma_i$ が $\mathbf{Y}$ の分布のパラメータであることを示す．

(vii) $\mathbf{U}_i$，$i=1,\cdots,n$ は $N_p(\boldsymbol{\mu},\Sigma)$ という独立な同一分布に従う．すなわち，i.i.d. (independent and identically distributed) とすると，

$$\frac{1}{n}\sum\mathbf{U}_i=\overline{\mathbf{U}}\sim N_p\left(\boldsymbol{\mu},\frac{1}{n}\Sigma\right). \tag{8a. 2. 19}$$

(8a. 2. 18) から導かれる (8a. 2. 19) は n 個の独立な測定値の平均値の分布を与える [(8a. 2. 18) で $A_i=1/n$ とせよ]．

(viii) $\mathbf{U}\sim N_p(\boldsymbol{\mu},\Sigma)$ とすると，

$$Q=(\mathbf{U}-\boldsymbol{\mu})'\mathbf{A}(\mathbf{U}-\boldsymbol{\mu})\sim\chi^2(k)$$

となるための必要十分条件は $\Sigma(\mathbf{A}\Sigma\mathbf{A}-\mathbf{A})\Sigma=0$ である．ここで $k=\mathrm{trace}(\mathbf{A}\Sigma)$．

(8a. 2. 9) から，$\mathbf{U}-\boldsymbol{\mu}=\mathbf{BG}$，ただし $\mathbf{G}\sim N_m(0,\mathbf{I})$ かつ $\Sigma=\mathbf{BB}'$．

$$Q=(\mathbf{U}-\boldsymbol{\mu})'\mathbf{A}(\mathbf{U}-\boldsymbol{\mu})=\mathbf{G}'\mathbf{B}'\mathbf{ABG} \tag{8a. 2. 20}$$

したがって，Q は独立正規変量 G_i を含む2次形式であらわされる．独立変量では2次形式 Q が χ^2 分布をもつための必要十分条件は Q の行列がベキ等であることである [3b. 4の(ii)]．したがって，

$$(\mathbf{B}'\mathbf{AB})(\mathbf{B}'\mathbf{AB})=\mathbf{B}'\mathbf{AB},\quad \text{すなわち}\quad \mathbf{B}'(\mathbf{A}\Sigma\mathbf{A}-\mathbf{A})\mathbf{B}=0.$$

しかし $\mathbf{B}'(\mathbf{A}\Sigma\mathbf{A}-\mathbf{A})\mathbf{B}=0\Leftrightarrow\mathbf{BB}'(\mathbf{A}\Sigma\mathbf{A}-\mathbf{A})\mathbf{BB}'=0=\Sigma(\mathbf{A}\Sigma\mathbf{A}-\mathbf{A})\Sigma$．$\chi^2$ の自由度は (1変量理論より)

$$\mathrm{rank}(\mathbf{B}'\mathbf{AB})=\mathrm{trace}(\mathbf{B}'\mathbf{AB})=\mathrm{trace}(\mathbf{ABB}')=\mathrm{trace}(\mathbf{A}\Sigma).$$

もし Σ が正則であれば，必要十分条件は $\mathbf{A}\Sigma\mathbf{A}=\mathbf{A}$ である．

8a. 3　N_p のいくつかの特徴づけ

(i) N_p の特徴づけ1　$\mathbf{U}_1$ と $\mathbf{U}_2$ を

$$\mathbf{U}=\mathbf{U}_1+\mathbf{U}_2\sim N_p$$

となる2つの独立な p 次元確率変数とすれば，$\mathbf{U}_1$ と $\mathbf{U}_2$ はいずれも N_p である．

仮定に従って，$\mathbf{L}'\mathbf{U}=\mathbf{L}'\mathbf{U}_1+\mathbf{L}'\mathbf{U}_2\sim N_1$ を考える．ただし $\mathbf{L}'\mathbf{U}_1$ と $\mathbf{L}'\mathbf{U}_2$ は独立である．したがって Cramér (1937) の定理により，$\mathbf{L}'\mathbf{U}_1$ と $\mathbf{L}'\mathbf{U}_2$ はそれぞれ N_1 である．$\mathbf{L}$ は任意であるので，$\mathbf{U}_1$ と $\mathbf{U}_2$ はそれぞれ N_p である．

(ii) N_p の特徴づけ2　n 個の独立な p 次元確率変数 $\mathbf{U}_1,\cdots,\mathbf{U}_n$ の2つの線形関数を考える．

$$\mathbf{V}_1=B_1\mathbf{U}_1+\cdots+B_n\mathbf{U}_n,\qquad \mathbf{V}_2=C_1\mathbf{U}_1+\cdots+C_n\mathbf{U}_n$$

$\mathbf{B}'=(B_1,\cdots,B_n)$ および $\mathbf{C}'=(C_1,\cdots,C_n)$ とすると，

(a) $\mathbf{U}_i$ は i. i. d. で N_p に従い，かつ $\mathbf{B}'\mathbf{C}=0$ であれば，$\mathbf{V}_1$ と $\mathbf{V}_2$ は独立に分布する．

(b) $\mathbf{V}_1$ と $\mathbf{V}_2$ が独立に分布するならば，$B_iC_i\neq0$ となる任意の i で $\mathbf{U}_i\sim N_p$ である．このとき，$\mathbf{U}_i$ が同一分布である必要はない．

(a) を証明するために，$\mathbf{V}_1,\mathbf{V}_2$ の同時分布が N_{2p} であることを定義1から導く．したがって独立性を証明するためには，$\mathrm{cov}(\mathbf{V}_1,\mathbf{V}_2)=0$ を示すだけでよい．

$$\mathrm{cov}(\mathbf{V}_1,\mathbf{V}_2)=\mathrm{cov}(B_1\mathbf{U}_1+\cdots+B_n\mathbf{U}_n,C_1\mathbf{U}_1+\cdots+C_n\mathbf{U}_n)$$
$$=B_1C_1\mathrm{cov}(\mathbf{U}_1,\mathbf{U}_1)+\cdots+B_nC_n\mathrm{cov}(\mathbf{U}_n,\mathbf{U}_n)$$
$$=(B_1C_1+\cdots+B_nC_n)\Sigma=0.$$

(b) の結果は1変量におけるダルモア＝スキトヴィッチの定理の p 次元確率変数への拡張である（第3章，問題 **21**）．ダルモア＝スキトヴィッチの定理によれば，$u_1, \cdots, u_n$ が独立な1次元確率変数であるとき，固定した B_i, C_i にたいし

$$B_1u_1 + \cdots + B_nu_n \text{ と } C_1u_1 + \cdots + C_nu_n$$

が独立であれば，$i = 1, \cdots, n$ のうち，$B_iC_i \neq 0$ なる i については，$u_i \sim N_1$ であり，それ以外の u_i の分布は任意で良い．この定理を用いれば，$\mathbf{V}_1, \mathbf{V}_2$ の線形関数

$$\mathbf{L}'\mathbf{V}_1 = \mathbf{L}'(B_1\mathbf{U}_1 + \cdots + B_n\mathbf{U}_n) = B_1(\mathbf{L}'\mathbf{U}_1) + \cdots + B_n(\mathbf{L}'\mathbf{U}_n),$$
$$\mathbf{L}'\mathbf{V}_2 = \mathbf{L}'(C_1\mathbf{U}_1 + \cdots + C_n\mathbf{U}_n) = C_1(\mathbf{L}'\mathbf{U}_1) + \cdots + C_n(\mathbf{L}'\mathbf{U}_n)$$

を考えるだけでよい．この $\mathbf{L}'\mathbf{U}_i$ はすべて1次元独立確率変数である．よって，$\mathbf{L}'\mathbf{U}_i$ は N_1 であるので，$\mathbf{U}_i$ は N_p に従う．

N_p の特徴づけ 3　N_p の定義2に従って，$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{BG}$ とする．ただし $\mathbf{G} \sim N_m(0, \mathbf{I})$，すなわち，$\mathbf{G}$ の成分は独立で，それぞれ $N_1(0, 1)$ の分布をする．しかし $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{BG}$ という表現は一意ではない．もし $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{B}_1\mathbf{G}_1$ と $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{B}_2\mathbf{G}_2$ が2つの異なる表現であれば，$D(\mathbf{Y}) = \mathbf{B}_1\mathbf{B}_1' = \mathbf{B}_2\mathbf{B}_2' = \boldsymbol{\Sigma}$ である．$\mathbf{Y}$ の表現が一意でないことを N_p の特徴づけに用いる．

(iii)　$\mathbf{Y} = \boldsymbol{\mu}_1 + \mathbf{B}_1\mathbf{G}_1$ および $\mathbf{Y} = \boldsymbol{\mu}_2 + \mathbf{B}_2\mathbf{G}_2$ を p 次元確率変数 $\mathbf{Y}$ の，非退化で各成分が独立な確率変数ベクトル（1変量正規の必要はない）$\mathbf{G}_1, \mathbf{G}_2$ による2つの表現とする．ただし，$\mathbf{B}_1$ および $\mathbf{B}_2$ は，階数 m の $p \times m$ 行列で $\mathbf{B}_1$ のどの列も $\mathbf{B}_2$ のどれかの列の定数倍になっていないとする．このとき $\mathbf{Y}$ は p 次元正規確率変数である．

$m < p$ の場合に $\mathbf{B}_1$ の列に直交する任意の p 次元ベクトルを $\mathbf{C}$ とすれば

$$\mathbf{C}'\mathbf{Y} = \mathbf{C}'\boldsymbol{\mu}_1 = \mathbf{C}'\boldsymbol{\mu}_2 + \mathbf{C}'\mathbf{B}_2\mathbf{G}_2$$

となり，$\mathbf{C}'\mathbf{B}_2\mathbf{G}_2$ は $\mathbf{C}'\mathbf{B}_2 = 0$ でなければ仮定に反して退化確率変数となることが導かれる．したがって，$\mathbf{B}_1$ と $\mathbf{B}_2$ の列は同じベクトル空間に属し，次数および階数 m の正方行列 $\mathbf{H}$ が存在して，$\mathbf{B}_1 = \mathbf{B}_2\mathbf{H}$ と書ける．$m = p$ の場合 $\mathbf{B}_1 = \mathbf{B}_2\mathbf{H}$ と書けることは自明である．次に

$$\mathbf{Y} - \boldsymbol{\mu}_2 = \mathbf{B}_2\mathbf{G}_2 \Rightarrow (\mathbf{B}_2'\mathbf{B}_2)^{-1}\mathbf{B}_2'(\mathbf{Y} - \boldsymbol{\mu}_2) = \mathbf{G}_2$$
$$\mathbf{Y} - \boldsymbol{\mu}_1 = \mathbf{B}_1\mathbf{G}_1 = \mathbf{B}_2\mathbf{H}\mathbf{G}_1 \Rightarrow (\mathbf{B}_2'\mathbf{B}_2)^{-1}\mathbf{B}_2'(\mathbf{Y} - \boldsymbol{\mu}_1) = \mathbf{H}\mathbf{G}_1$$

よって，$\mathbf{G}_2$ と $\mathbf{H}\mathbf{G}_1$ は平均値は別として同一分布をする．したがって，$\mathbf{H}\mathbf{G}_1$ の成分は独立に分布する．以上の結果 $\mathbf{G}_1 \Rightarrow \mathbf{H}\mathbf{G}_1$ の変換は確率変数の独立性を保持している．

$\mathbf{B}_1$ のどの列ベクトルも $\mathbf{B}_2$ の1つの列ベクトルではあらわせないという条件より，$\mathbf{H}$ に $m-1$ 個の0をもつ列がないことがわかる．このことは，共通する行がないように $\mathbf{H}$ の行の組を2つ選び，その1次結合をつくって，その i 番目と j 番目の係数がともに0にならないようにできることと同じである．よって独立に分布するような $\mathbf{G}_1$ の成分の2つの1次結合を得たことになり，(ii) で述べたダルモア＝スキトヴィッチの定理により，i 番目と j 番目の成分は正規分布をする．i と j は任意であるので，求める結果は証明された．

N_p の他の特徴づけに関しては，Kagan, Linnik and Rao (1972B4) による著書と，Khatri and Rao (1972f) および Rao (1966a, 1967d, 1969a, b) の論文を参照せよ．

8a.4　多変量正規分布の密度関数

本書では，**8a.1** および **8a.2** における議論で多変量正規分布の密度関数を用いなかった（つまり密度関数の存在を証明しなかった）．どんな形で密度関数を決定できるであろうかという問の答は，分散共分散行列 $\boldsymbol{\Sigma}$ の階数 $R(\boldsymbol{\Sigma})$ によって変わる．

場合 1.　$R(\boldsymbol{\Sigma}) = p$．$\mathbf{U} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ とすれば，(8a.2.9) に示したように，

$$\mathbf{U} - \boldsymbol{\mu} = \mathbf{BG}, \quad \text{ただし} \quad \mathbf{G} \sim N_m(0, \mathbf{I}), \ \mathbf{BB}' = \boldsymbol{\Sigma} \tag{8a.4.1}$$

であり，$m = \operatorname{rank} \boldsymbol{\Sigma} = p$ である．したがって $|\boldsymbol{\Sigma}| = |\mathbf{BB}'| \neq 0$ すなわち $|\mathbf{B}| \neq 0$ であり，このとき次の反転関係が存在する．

$$\mathbf{G} = \mathbf{B}^{-1}(\mathbf{U} - \boldsymbol{\mu}). \tag{8a.4.2}$$

$\mathbf{G} \sim N_p(0, \mathbf{I})$ であるので，$\mathbf{G}$ の密度は

$$(2\pi)^{-p/2}\exp\left[-\frac{1}{2}\mathbf{G}'\mathbf{G}\right]. \tag{8a.4.3}$$

(8a.4.2) の関係を用いて，$\mathbf{U} - \boldsymbol{\mu}$ で書きなおすと，ヤコビアンは $|\mathbf{B}|^{-1} = |\boldsymbol{\Sigma}|^{-1/2}$ ($\mathbf{BB}' = \boldsymbol{\Sigma}$ であるので) であり，密度 (8a.4.3) は

$$(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{U} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{U} - \boldsymbol{\mu})\right]. \tag{8a.4.4}$$

となる．これは R^p におけるルベーグ測度に関する $\mathbf{U}$ の密度である．

場合 2.　$R(\boldsymbol{\Sigma}) = k < p$．$R(\boldsymbol{\Sigma}) < p$ の場合は，反転 (8a.4.2) は不可能で，したがって R^p におけるルベーグ測度に関する密度関数の陽な表現は不可能である．しかし，部分空間における密度関数は存在する．

$\mathbf{B}$ を $\mathscr{M}(\boldsymbol{\Sigma})$ に属する正規直交ベクトルのつくる $p \times k$ 行列とし，$\mathbf{N}$ を $\mathbf{N}'\boldsymbol{\Sigma} = 0$ となるような階数 $p-k$ をもつ $p \times (p-k)$ 行列とする．$\mathbf{X} = \mathbf{B}'\mathbf{U}$，$\mathbf{Z} = \mathbf{N}'\mathbf{U}$ として，$\mathbf{U}' \to (\mathbf{X}' \,\vdots\, \mathbf{Z}')$ なる変換を考える．そうすれば，

$$E(\mathbf{Z}) = \mathbf{N}'\boldsymbol{\mu}, \qquad D(\mathbf{Z}) = \mathbf{N}'\boldsymbol{\Sigma}\mathbf{N} = 0 \tag{8a.4.5}$$

したがって

$$\mathbf{Z} = \mathbf{N}'\boldsymbol{\mu} \quad \text{確率 1 で}. \tag{8a.4.6}$$

さらに

$$E(\mathbf{X}) = \mathbf{B}'\boldsymbol{\mu}, \qquad D(\mathbf{X}) = \mathbf{B}'\boldsymbol{\Sigma}\mathbf{B}, \tag{8a.4.7}$$

したがって

$$\mathbf{X} \sim N_k(\mathbf{B}'\boldsymbol{\mu}, \mathbf{B}'\Sigma\mathbf{B}). \tag{8a.4.8}$$

$|\mathbf{B}'\Sigma\mathbf{B}|$ は Σ の零でない固有値の積 $\lambda_1\cdot\cdots\cdot\lambda_k$ とみなせるので，$\mathbf{B}'\Sigma\mathbf{B}$ は特異でない．この場合 $\mathbf{X}$ は密度

$$\frac{(2\pi)^{-k/2}}{|\mathbf{B}'\Sigma\mathbf{B}|^{1/2}}e^{-(\mathbf{X}-\mathbf{B}'\boldsymbol{\mu})'(\mathbf{B}'\Sigma\mathbf{B})^{-1}(\mathbf{X}-\mathbf{B}'\boldsymbol{\mu})/2} \tag{8a.4.9}$$

をもつ．こうして $\mathbf{U}$ の分布，すなわち，それと等価な $\mathbf{X}, \mathbf{Z}$ の分布は，(8a.4.6) と (8a.4.9) によって規定される．

一般逆行列 Σ^- を任意に選んでも

$$\begin{aligned}(\mathbf{X}-\mathbf{B}'\boldsymbol{\mu})'(\mathbf{B}'\Sigma\mathbf{B})^{-1}(\mathbf{X}-\mathbf{B}'\boldsymbol{\mu}) &= (\mathbf{U}-\boldsymbol{\mu})'\mathbf{B}(\mathbf{B}'\Sigma\mathbf{B})^{-1}\mathbf{B}'(\mathbf{U}-\boldsymbol{\mu}) \\ &= (\mathbf{U}-\boldsymbol{\mu})'\Sigma^-(\mathbf{U}-\boldsymbol{\mu}) \end{aligned} \tag{8a.4.10}$$

が成り立つことがわかる．それゆえ，密度 (8a.4.9) は

$$\frac{(2\pi)^{-k/2}}{\sqrt{\lambda_1\cdot\cdots\cdot\lambda_k}}e^{-(\mathbf{U}-\boldsymbol{\mu})'\Sigma^-(\mathbf{U}-\boldsymbol{\mu})/2} \tag{8a.4.11}$$

と書くことができ，一方 (8a.4.6) は

$$\mathbf{N}'\mathbf{U} = \mathbf{N}'\boldsymbol{\mu} \quad \text{確率 1 で} \tag{8a.4.12}$$

と書くことができる．こうして $\mathbf{U}$ の分布は (8a.4.11) と (8a.4.12) で規定される．ここで前者は超平面 $\mathbf{N}'\mathbf{U} = \mathbf{N}'\boldsymbol{\mu}$ 上の密度と解釈される．(以上の表現は Khatri (1968) による.)

8a.5　母数の推定

n 個の p 変量母集団とそれぞれよりの独立な観測値を考える．$\mathbf{U}_i$ を i 番目の p 変量母集団よりの観測値とする．それらの観測値は行列の形で書くことができる．

$$\begin{array}{c|cccc|} & \mathbf{U}_1 & \mathbf{U}_2 & & \mathbf{U}_n \\ \mathbf{Y}_1' & U_{11} & U_{12} & \cdots & U_{1n} \\ & \cdot & \cdot & \cdots & \cdot \\ \mathbf{Y}_p' & U_{p1} & U_{p2} & \cdots & U_{pn} \end{array} = \mathscr{U}' \tag{8a.5.1}$$

ただし，i 番目の列ベクトル $\mathbf{Y}_i'$ は p 次元変数の i 番目の成分についての n 個の独立な観測値を示す．$\mathscr{U}$ は次数 $n \times p$ の全体の行列を示す．本章では最後までこの記号(8a.5.1)を用いる．

i.i.d. 変数に基づく $\boldsymbol{\mu}$ と Σ の推定　$\mathbf{U}_1, \cdots, \mathbf{U}_n$ は独立で，平均 $\boldsymbol{\mu}$ と分散共分散行列 Σ の同一分布に従うとすると，

$$U_{i1}, \cdots, U_{in}$$

は平均 μ_i，分散 σ_{ii} の 1 変量分布からの i.i.d. 観測値である．それゆえ 1 変量の計算から

$$\bar{Y}_i = \bar{U}_i = \frac{1}{n}\sum_j U_{ij} \quad \text{は } \mu_i \text{ にたいして不偏であり}$$

$$S_{ii} = \sum_j (U_{ij} - \bar{U}_i)^2 \quad \text{は } (n-1)\sigma_{ii} \text{ にたいして不偏であり}$$

$$s_{ii} = \frac{S_{ii}}{n-1} \quad \text{は } \sigma_{ii} \text{ にたいして不偏である．}$$

いま，

$$U_{i1} + U_{j1}, \quad \cdots, \quad U_{in} + U_{jn}$$

を考えると，それらは平均 $\mu_i + \mu_j$ と分散 $\sigma_{ii} + 2\sigma_{ij} + \sigma_{jj}$ をもつ 1 変量分布からの i.i.d. 観測値である．したがって，

$$\sum (U_{ir} + U_{jr} - \bar{U}_i - \bar{U}_j)^2 = S_{ii} + S_{jj} + 2S_{ij}$$

は $(n-1)(\sigma_{ii} + 2\sigma_{ij} + \sigma_{jj})$ にたいして不偏である．ただし

$$S_{ij} = \sum (U_{ir} - \bar{U}_i)(U_{jr} - \bar{U}_j).$$

したがって，$S_{ii}/(n-1)$ と $S_{jj}/(n-1)$ がそれぞれ σ_{ii} と σ_{jj} にたいして不偏であるので，$S_{ij}/(n-1)$ は σ_{ij} にたいして不偏となる．

こうして，$\mathbf{U}_i$ の分布の正規性を仮定しない一般の場合について，$\boldsymbol{\mu}'$ と Σ の不偏推定量として

$$\bar{\mathbf{U}}' = (\bar{U}_1, \cdots, \bar{U}_p). \qquad \frac{1}{n-1}(S_{ij}) = \frac{1}{n-1}\mathbf{S}$$

が得られる．記号 (8a.5.1) を使って，(S_{ij}) の別の行列表現を与えることは興味あることである．この証明は読者ができるだろう．

$$S_{ij} = \mathbf{Y}_i'\mathbf{Y}_j - n\bar{U}_i\bar{U}_j$$

$$\begin{aligned}\mathbf{S} = (S_{ij}) &= \sum_1^n (\mathbf{U}_r - \bar{\mathbf{U}})(\mathbf{U}_r - \bar{\mathbf{U}})' = \sum_1^n \mathbf{U}_r\mathbf{U}_r' - n\bar{\mathbf{U}}\bar{\mathbf{U}}' \\ &= \mathscr{U}'\mathscr{U} - n\bar{\mathbf{U}}\bar{\mathbf{U}}'\end{aligned} \tag{8a.5.2}$$

最尤法による推定　$R(\Sigma) = p$ がわかっている場合，ひとつの観測値 $\mathbf{U}$ の密度関数は，(8a.4.4) に示したように，

$$(2\pi)^{-p/2}|(\sigma^{ij})|^{1/2}\exp\left[-\frac{1}{2}\operatorname{trace}\{\Sigma^{-1}(\mathbf{U}-\boldsymbol{\mu})(\mathbf{U}-\boldsymbol{\mu})'\}\right]$$

である．ただし，$(\sigma^{ij}) = (\sigma_{ij})^{-1} = \Sigma^{-1}$．$n$ 個の観側値 $(\mathbf{U}_1, \cdots, \mathbf{U}_n)$ の同時分布の密度関数は，定数を除けば，

$$|(\sigma^{ij})|^{n/2}\exp\left[-\frac{1}{2}\operatorname{trace}\left\{\Sigma^{-1}\sum_1^n(\mathbf{U}_r-\boldsymbol{\mu})(\mathbf{U}_r-\boldsymbol{\mu})'\right\}\right] \tag{8a.5.3}$$

であらわせる．ここで，

$$\sum_1^n (\mathbf{U}_r - \boldsymbol{\mu})(\mathbf{U}_r - \boldsymbol{\mu})' = \sum_1^n (\mathbf{U}_r - \bar{\mathbf{U}})(\mathbf{U}_r - \bar{\mathbf{U}})' + n(\bar{\mathbf{U}} - \boldsymbol{\mu})(\bar{\mathbf{U}} - \boldsymbol{\mu})'$$

$$= \mathbf{S} + n(\bar{\mathbf{U}} - \boldsymbol{\mu})(\bar{\mathbf{U}} - \boldsymbol{\mu})'$$

に注意すれば，

$$\mathrm{trace}\left\{\boldsymbol{\Sigma}^{-1}\sum_{1}^{n}(\mathbf{U}_r - \boldsymbol{\mu})(\mathbf{U}_r - \boldsymbol{\mu})'\right\}$$
$$= \mathrm{trace}\,\{\boldsymbol{\Sigma}^{-1}\mathbf{S}\} + n\,\mathrm{trace}\,\{\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{U}} - \boldsymbol{\mu})(\bar{\mathbf{U}} - \boldsymbol{\mu})'\}$$
$$= \sum\sum \sigma^{ij}S_{ij} + n(\bar{\mathbf{U}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{U}} - \boldsymbol{\mu})$$

であることがわかる．したがって (8a. 5. 3) の対数は

$$\frac{n}{2}\log|(\sigma^{ij})| - \frac{1}{2}\sum\sum\sigma^{ij}S_{ij} - \frac{n}{2}(\bar{\mathbf{U}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{U}} - \boldsymbol{\mu}) \qquad (8a.\ 5.\ 4)$$

または，別の形では

$$\frac{n}{2}\log|(\sigma^{ij})| - \frac{1}{2}\sum\sum\sigma^{ij}[S_{ij} + n(\bar{U}_i - \mu_i)(\bar{U}_j - \mu_j)] \qquad (8a.\ 5.\ 5)$$

と書くことができる．(8a. 5. 4) の形式を用いて，$\boldsymbol{\mu}$ について微分すれば

$$\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{U}} - \boldsymbol{\mu}) = 0 \quad \text{すなわち，} \quad \bar{\mathbf{U}} - \boldsymbol{\mu} = 0$$

が得られる．これより推定量 $\hat{\boldsymbol{\mu}} = \bar{\mathbf{U}}$ すなわち，$\hat{\mu}_i = \bar{U}_i$, $i = 1, \cdots, p$ が導かれる．

第2の形式 (8a. 5. 5) は σ^{ij} について微分するのに都合がよい．それより

$$\frac{n}{|(\sigma^{ij})|}\frac{\partial|(\sigma^{ij})|}{\partial\sigma^{ij}} = [S_{ij} + n(\bar{U}_i - \mu_i)(\bar{U}_j - \mu_j)] \qquad (8a.\ 5.\ 6)$$

が得られる．

$$\frac{\partial|(\sigma^{ij})|}{\partial\sigma^{ij}} = (\sigma^{ij}) \text{ における } \sigma^{ij} \text{ の余因子}$$

であるから，

$$\frac{n}{|(\sigma^{ij})|}\frac{\partial|(\sigma^{ij})|}{\partial\sigma^{ij}} = n\sigma_{ij}$$

となることが容易にわかる．さらに $(\bar{U}_i - \mu_i) = 0$ を代入すると，式 (8a. 5. 6) は

$$n\sigma_{ij} - S_{ij} = 0$$

となり，σ_{ij} の推定量として

$$\hat{\sigma}_{ij} = \frac{S_{ij}}{n}. \qquad E(\hat{\sigma}_{ij}) = \frac{E(S_{ij})}{n} = \frac{n-1}{n}\sigma_{ij}$$

が得られ，これは少し偏りがあることがわかる．このとき σ^{ij} の推定量は nS^{ij} である．ただし，$(S^{ij}) = (S_{ij})^{-1}$.

母数を推定量でおきかえたときの対数尤度 (8a. 5. 4) の値は次のようになる．

$$\frac{n}{2}\log n^p + \frac{n}{2}\log|(S^{ij})| - \frac{n}{2}\sum\sum S^{ij}S_{ij} = \frac{n}{2}\log n^p - \frac{n}{2}\log|(S_{ij})| - \frac{np}{2}. \qquad (8a.\ 5.\ 7)$$

ここで，対数尤度の導関数を0とおくことによって得られた推定量 $\hat{\boldsymbol{\mu}} = \bar{\mathbf{U}}$ および $(\hat{\sigma}_{ij}) = n^{-1}(S_{ij}) = n^{-1}\mathbf{S}$ が，事実，最尤推定量になっていることを証明しよう．

それには，すべての $\boldsymbol{\mu}$ および $\boldsymbol{\Sigma}$ にたいして (8a. 5. 5) ⩽ (8a. 5. 7) を証明すれば十分である．その差 (8a. 5. 7) − (8a. 5. 5) は

$$-\frac{n}{2}\log\frac{|n^{-1}\mathbf{S}|}{|\boldsymbol{\Sigma}|} - \frac{np}{2} + \frac{1}{2}\mathrm{trace}\,\{\boldsymbol{\Sigma}^{-1}[\mathbf{S} + n(\bar{\mathbf{U}} - \boldsymbol{\mu})(\bar{\mathbf{U}} - \boldsymbol{\mu})']\}$$

に等しく，$\lambda_1, \cdots, \lambda_p$ を固有方程式 $|n^{-1}\mathbf{S} - \lambda\boldsymbol{\Sigma}| = 0$ の根とすると，

$$\geqslant -\frac{n}{2}\log\frac{|n^{-1}\mathbf{S}|}{|\boldsymbol{\Sigma}|} - \frac{np}{2} + \frac{1}{2}\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})$$
$$= \frac{n}{2}[-\log(\lambda_1 \cdots \lambda_p) - p + (\lambda_1 + \cdots + \lambda_p)] \qquad (8a.\ 5.\ 8)$$

となる．ところで，任意の非負の x にたいして，$x \leqslant e^{x-1}$ であるから，対数をとると

$$-\log x - 1 + x \geqslant 0 \qquad (8a.\ 5.\ 9)$$

である．(8a. 5. 8) の大括弧の内側の値は，

$$\sum_{i=1}^{p}(-\log\lambda_i - 1 + \lambda_i) \qquad (8a.\ 5.\ 10)$$

となり，この各項へ不等式 (8a. 5. 9) を適用すると，これは $\geqslant 0$ であることがわかる．このエレガントな証明は Watson (1964) によるものである．

対数尤度はこれらの推定量と母数のみの関数としてあらわせるから，この最尤推定量 $\bar{\mathbf{U}}$ と $n^{-1}\mathbf{S}$ は母数 $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ にたいする，十分統計量でもあることが証明できる．

$R(\boldsymbol{\Sigma})$ が未知であれば，それが p より小さい可能性も考慮に入れねばならない．この場合には，適当な超平面 (8a. 4. 12) または (8a. 4. 6) を定め，密度関数 (8a. 4. 11) または (8a. 4. 9) を用いねばならない．

密度 (8a. 4. 9) を採用し，すでに証明した結果を用いると，$\mathbf{B}'\boldsymbol{\mu}$ および $\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B}$ の最尤推定量は

$$\mathbf{B}'\bar{\mathbf{U}} = \mathbf{B}'\hat{\boldsymbol{\mu}}, \qquad \mathbf{B}'\mathbf{S}\mathbf{B} = n\,\mathbf{B}'\hat{\boldsymbol{\Sigma}}\mathbf{B}$$

を満足し，一方 (8a. 4. 6) は

$$\mathbf{N}'\bar{\mathbf{U}} = \mathbf{N}'\boldsymbol{\mu}, \qquad \mathbf{N}'\mathbf{S}\mathbf{N} = n\,\mathbf{N}'\boldsymbol{\Sigma}\mathbf{N} = 0$$

を与えるから，結局 $\boldsymbol{\Sigma}$ が正則の場合と同じ次の解が得られる．

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{U}} \quad \text{と} \quad \hat{\boldsymbol{\Sigma}} = n^{-1}\mathbf{S}.$$

8a. 6　最大エントロピーをもつ分布としての N_p

3a. 1 で，平均と分散を固定した場合に，1変量正規確率密度は最大エントロピーをもつことを示した．同じ結果が N_p についても成り立つ．

$P(\mathbf{U})$ を p 次元確率変数 $\mathbf{U}$ の確率密度とする．問題は，dv を体積要素とするとき，

$$-\int P(\mathbf{U})\log P(\mathbf{U})dv \tag{8a.6.1}$$

を，平均と分散共分散行列が一定という条件の下で最大にすることである．つまり，

$$\left.\begin{array}{l}\int P(\mathbf{U})dv=1, \qquad \int \mathbf{U}P(\mathbf{U})dv=\mu \\ \int[(\mathbf{U}-\mu)(\mathbf{U}-\mu)']P(\mathbf{U})dv=\Sigma\end{array}\right\}. \tag{8a.6.2}$$

条件 (8a.6.2) を満足するように正規確率密度

$$Q(\mathbf{U})=(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{U}-\mu)'\Sigma^{-1}(\mathbf{U}-\mu)\right], \tag{8a.6.3}$$

を選ぶ．情報理論の不等式 (1e.6.6) から，任意の2つの密度関数 $P(\mathbf{U}), Q(\mathbf{U})$ にたいして

$$-\int P\log P dv \leqslant -\int P\log Q\, dv$$

が成り立つ．ここで等号は(dv) のほとんどいたるところで $P=Q$ のときに成立する．このQ に (8a.6.3) を代入し，また (8a.6.2) を使うと，

$$\begin{aligned}-\int P\log P dv &\leqslant -\int P(\mathbf{U})\left[-\frac{p}{2}\log(2\pi)-\frac{1}{2}\log|\Sigma|-\frac{1}{2}(\mathbf{U}-\mu)'\Sigma^{-1}(\mathbf{U}-\mu)\right]dv \\ &=\frac{p}{2}\log(2\pi)+\frac{1}{2}\log|\Sigma|+\frac{1}{2}\operatorname{trace}\left\{\Sigma^{-1}\int[(\mathbf{U}-\mu)(\mathbf{U}-\mu)']P(\mathbf{U})dv\right\} \\ &=\frac{p}{2}\log(2\pi)+\frac{1}{2}\log|\Sigma|+\frac{1}{2}\operatorname{trace}\Sigma^{-1}\Sigma \end{aligned} \tag{8a.6.4}$$

となる．式 (8a.6.4) はエントロピーに一定の上限があることを示している．その上限は正規確率密度 (8a.6.3) によって達成され，その値は次のようになる．

$$\frac{p}{2}\log(2\pi)+\frac{p}{2}+\frac{1}{2}\log|\Sigma|.$$

8b　ウィシャート分布

8b.1　定 義 と 記 号

$\mathbf{U}_i \sim N_p(\mu_i, \Sigma)$, $i=1,\cdots,k$ はすべて互いに独立とし，$\mathbf{Y}_j, \mathscr{U}$ は (8a.5.1) の n の代りに k を用いて定義する．なお，$\mathbf{M}'$ は $\mu_1,\cdots,\mu_k$ を列とする $p\times k$ 行列とする．すなわち

$$\begin{array}{c|ccc|l} & \mathbf{U}_1 & \cdots & \mathbf{U}_k & \\ \hline \mathbf{Y}_1' & U_{11} & \cdots & U_{1k} & \\ \vdots & \vdots & \cdots & \vdots & =\mathscr{U}' \\ \mathbf{Y}_p' & U_{p1} & \cdots & U_{pk} & \\ \hline \mathbf{Y}' & \mathbf{L}'\mathbf{U}_1 & \cdots & \mathbf{L}'\mathbf{U}_k & =\mathbf{L}'\mathscr{U}' \end{array} \tag{8b.1.1}$$

である．ただし $\mathbf{L}$ は $L_1,\cdots,L_p$ を要素とする定数ベクトルである．

以下の議論では，各 i について $\mathbf{U}_i$ の線型関数 $\mathbf{L}'\mathbf{U}_i$ をしばしば引用する．ここで，$E(\mathbf{L}'\mathbf{U}_i)=\mathbf{L}'\mu_i$ かつ，$V(\mathbf{L}'\mathbf{U}_i)=\mathbf{L}'\Sigma\mathbf{L}$ であるので，

$$\mathbf{L}'\mathbf{U}_i \sim N_1(\mathbf{L}'\mu_i, \sigma_L^2), \qquad \sigma_L^2=\mathbf{L}'\Sigma\mathbf{L}, \quad i=1,\cdots,k \tag{8b.1.2}$$

と書け，また，$\mathbf{U}_i$ が独立であるので，$\mathbf{L}'\mathbf{U}_i$ はすべて独立である．k 個の独立な N_1 確率変数のベクトル $\mathbf{Y}'=(\mathbf{L}'\mathbf{U}_1,\cdots,\mathbf{L}'\mathbf{U}_k)$ は

$$\mathbf{Y}=L_1\mathbf{Y}_1+\cdots+L_p\mathbf{Y}_p=\mathscr{U}\mathbf{L} \tag{8b.1.3}$$

と表現できる．$\mathbf{Y}=\mathscr{U}\mathbf{L}$ の分布は同一の分散 σ_L^2 をもつ k 個の独立な N_1 確率変数の分布であるが，平均 $\mathbf{L}'\mu_1,\cdots,\mathbf{L}'\mu_k$ は異なる．すなわち

$$\mathbf{Y}\sim N_k(\mathbf{ML}, \sigma_L^2\mathbf{I}) \tag{8b.1.4}$$

である．

さらにベクトル変数 $\mathbf{U}_i$ の1次結合を考えると，これは [(vi), **8a.2**] の再生性によって N_p の分布に従う．$\mathbf{B}'=(B_1,\cdots,B_k)$ を定数ベクトルとすると，(8a.2.18) の結果から，

$$B_1\mathbf{U}_1+\cdots+B_k\mathbf{U}_k=\mathscr{U}'\mathbf{B}\sim N_p(\mathbf{M}'\mathbf{B}, (\textstyle\sum B_i^2)\Sigma) \tag{8b.1.5}$$

となる．さらに，$\mathbf{B}_1'\mathbf{B}_2=0$ ならば $\mathscr{U}'\mathbf{B}_1$ と $\mathscr{U}'\mathbf{B}_2$ は互いに独立に分布することが [(ii), **8a.3**] で示されている．したがって，1次結合

$$\mathbf{V}_1=\mathscr{U}'\mathbf{B}_1, \ \cdots, \ \mathbf{V}_r=\mathscr{U}'\mathbf{B}_r, \qquad r\leqslant k \tag{8b.1.6}$$

は，ベクトル $\mathbf{B}_1,\cdots,\mathbf{B}_r$ が正規直交系のとき，同一の分散共分散行列 Σ をもつ N_p として，すべて互いに独立に分布することが導かれる．

$\mathbf{H}$ を k 次の直交行列とすると，変換

$$\mathscr{V}=\mathscr{U}'\mathbf{H} \tag{8b.1.7}$$

は，列ベクトル $\mathbf{U}_1,\cdots,\mathbf{U}_k$ を列ベクトル $\mathbf{V}_1,\cdots,\mathbf{V}_k$ へうつす．ただし，$\mathbf{V}_i$ は $\mathbf{U}_1,\cdots,\mathbf{U}_k$ の1次結合である．(8b.1.6) の結果は $\mathbf{V}_i$ が互いに独立に分布することを示している（1変量の場合と同様に，直交変換は独立性を保持する）．

Wishart (1928) 分布の定義　$\mathbf{U}_r, \mathbf{Y}_i$ および $\mathscr{U}$ と $\mathbf{M}$ が (8b.1.1) で定義されているとおりならば，

$$(S_{ij})=\mathbf{S}=\sum_1^k \mathbf{U}_r\mathbf{U}_r'=(\mathbf{Y}_i'\mathbf{Y}_j)=\mathscr{U}'\mathscr{U} \tag{8b.1.8}$$

という行列の要素の同時分布は自由度 k のウィシャート分布とよび，$W_p(k,\Sigma,\mathbf{M})$ であらわす．$\mathbf{M}=0$ のとき，この分布は中心ウィシャート分布といい，$W_p(k,\Sigma)$ と書く．われわれは中心または非心分布のいずれかをあらわすとき，記号 $W_p(k,\Sigma,\cdot)$ を用いる．$p=1$ のとき，$W_1(k,\sigma^2)$ は自由度 k の χ^2 分布，$\sigma^2\chi^2(k)$ と同じである．すなわち，ウィ

シャート分布は χ^2 分布の多変量への一般化にあたる．

N_p の場合と同様に，W_p の密度関数はつねに存在するとはかぎらない．それは本章末の問題 **11** で導くように，$k \geq p$ のときに存在する．われわれの研究では W_p の確率密度は必要としないであろう．ここでは W_p の定義から非常に一般的な結果を導く．

行列 $\mathbf{S} = (S_{ij})$ はより明確な形では

$$(S_{ij}) = \mathscr{U}'\mathscr{U} = \begin{pmatrix} \mathbf{Y}_1'\mathbf{Y}_1 & \mathbf{Y}_1'\mathbf{Y}_2 & \cdots & \mathbf{Y}_1'\mathbf{Y}_p \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{Y}_p'\mathbf{Y}_1 & \mathbf{Y}_p'\mathbf{Y}_2 & \cdots & \mathbf{Y}_p'\mathbf{Y}_p \end{pmatrix}$$

であり，i 番目の対角要素はベクトル変数 $\mathbf{U}_r$ の i 番目の成分の平方和であり，(i, j) 要素は i 成分と j 成分の積和である．$\mathscr{U}'\mathscr{U}$ という表現は1変量理論における平方和と似ている．さらに一般的な2次形式は行列 $\mathscr{U}'\mathbf{A}\mathscr{U}$

$$\mathscr{U}'\mathbf{A}\mathscr{U} = \begin{pmatrix} \mathbf{Y}_1'\mathbf{A}\mathbf{Y}_1 & \cdots & \mathbf{Y}_1'\mathbf{A}\mathbf{Y}_p \\ \cdots & \cdots & \cdots \\ \mathbf{Y}_p'\mathbf{A}\mathbf{Y}_1 & \cdots & \mathbf{Y}_p'\mathbf{A}\mathbf{Y}_p \end{pmatrix} \tag{8b.1.9}$$

であり，これは $\mathbf{Y}_1, \cdots, \mathbf{Y}_p$ の2次形式および双線形形式より成る．このとき，次のことが問題になる．$\mathscr{U}'\mathbf{A}\mathscr{U}$ は如何なる条件下でウィシャート分布に従うか？　その答は，任意の定数ベクトル $\mathbf{L}$ にたいして

$$\mathscr{U}'\mathbf{A}\mathscr{U} \sim W_p \Leftrightarrow \mathbf{L}'\mathscr{U}'\mathbf{A}\mathscr{U}\mathbf{L} \sim \chi^2 \tag{8b.1.10}$$

である．ここで，(8b.1.10) の右辺 $(\mathscr{U}\mathbf{L})'\mathbf{A}(\mathscr{U}\mathbf{L})$ は，k 個の独立な N_1 確率変数 $(\mathbf{L}'\mathbf{U}_1, \cdots, \mathbf{L}'\mathbf{U}_k)$ の2次形式 $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ である．独立な N_1 確率変数の2次形式の分布は，**3b.4** で詳細に研究したので，(8b.1.10) の結果から行列の分布への適当な一般化が可能になった．(8b.1.10) の証明は **8b.2** で行う．

8b.2　ウィシャート分布に関する若干の結果

（i）$S \sim W_p(k, \Sigma, \cdot)$ とし，$\mathbf{L}$ を任意の定数ベクトルとすれば，$\mathbf{L}'\mathbf{S}\mathbf{L} \sim \sigma_L^2\chi^2(k, \cdot)$ となり，W_p が中心分布であれば χ^2 も同様である．

定義より $\mathbf{S} = \sum \mathbf{U}_r\mathbf{U}_r'$ である．$\mathbf{L}'$ と $\mathbf{L}$ を前と後からかけると，$\mathbf{Y}$ は共通の分散 σ_L^2 をもつ独立な N_1 確率変数のベクトルである（(8b.1.4) 参照）ので，

$$\begin{aligned}\mathbf{L}'\mathbf{S}\mathbf{L} &= \sum (\mathbf{L}'\mathbf{U}_r)(\mathbf{U}_r'\mathbf{L}) = \sum (\mathbf{L}'\mathbf{U}_r)^2 \\ &= \mathbf{Y}'\mathbf{Y} \sim \sigma_L^2\chi^2(k, \cdot)\end{aligned}$$

となる．W_p が中心分布をするならば $\mathbf{M} = 0$ であり，この場合は $E(\mathbf{Y}) = \mathbf{ML} = 0$ となり，したがって χ^2 は中心分布である．

（i）の厳密な逆は真でないことが，Mitra (1970) によって示されていることに注意せよ．しかし，(ii) で与える逆の弱い表現は有用である．

（ii）$\mathbf{U}_i \sim N_p(\mu_i, \Sigma)$, $i = 1, \cdots, k$ は互いに独立で，$\mathscr{U}, \mathbf{Y}_i, \mathbf{Y}$ は (8b.1.1) で定義されたとおりとすると，$\mathscr{U}'\mathbf{A}\mathscr{U} \sim W_p(r, \Sigma, \cdot)$ であるための必要十分条件は，任意の定数ベクトル $\mathbf{L}$ にたいして，$\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \sigma_L^2\chi^2(r, \cdot)$ であることである，ただし $r = \text{rank}\,\mathbf{A} = \text{trace}\,\mathbf{A}$ である．また $\mathscr{U}'\mathbf{A}\mathscr{U} \sim W_p(r, \Sigma)$ であるための必要十分条件は任意の $\mathbf{L}$ について $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \sigma_L^2\chi^2(r)$ であることである．

必要性はより一般的な結果（i）の帰結である．十分性を証明するために（[(ii), **3b.4**] の1変量理論から），$\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \sigma_L^2\chi^2(r, \cdot)$ ならば $\mathbf{A}$ はベキ等で階数 r であることに注意しよう．すると，

$$\mathbf{A} = \mathbf{B}_1\mathbf{B}_1' + \cdots + \mathbf{B}_r\mathbf{B}_r' \tag{8b.2.1}$$

$$\begin{aligned}\mathscr{U}'\mathbf{A}\mathscr{U} &= \mathscr{U}'\mathbf{B}_1\mathbf{B}_1'\mathscr{U} + \cdots + \mathscr{U}'\mathbf{B}_r\mathbf{B}_r'\mathscr{U} \\ &= \mathbf{V}_1\mathbf{V}_1' + \cdots + \mathbf{V}_r\mathbf{V}_r'\end{aligned} \tag{8b.2.2}$$

となる r 個の正規直交ベクトル $\mathbf{B}_1, \cdots, \mathbf{B}_r$ が存在する．ただし，$\mathbf{V}_i = \mathscr{U}'\mathbf{B}_i$ である．$\mathbf{B}_i$ は正規直交系であるので，$\mathbf{V}_i$ は独立な N_p 確率変数 (8b.1.6) となる．したがって，十分性はウィシャート分布の定義から導かれる．

$\mathbf{Y} \sim N_k(\mathbf{ML}, \sigma_L^2\mathbf{I})$ であるので，$\mathbf{Y}'\mathbf{A}\mathbf{Y}$ の分布の非心母数は $\mathbf{Y} = \mathbf{ML}$ とおいたときの $\mathbf{Y}'\mathbf{A}\mathbf{Y}/\sigma_L^2$ の値，つまり $\mathbf{L}'\mathbf{M}'\mathbf{AML}/\sigma_L^2$ となる（(3b.2.2) 参照）．すべての $\mathbf{L}$ にたいして $\mathbf{Y}'\mathbf{A}\mathbf{Y}/\sigma_L^2 \sim \chi^2(r)$，つまり，中心 χ^2 分布に従うならば，

$$\frac{\mathbf{L}'\mathbf{M}'\mathbf{AML}}{\mathbf{L}'\Sigma\mathbf{L}} = 0, \quad \text{各 } \mathbf{L} \text{ たいして} \Rightarrow \mathbf{M}'\mathbf{AM} = 0$$

となる．一方 (8b.2.1) から

$$\begin{aligned}\mathbf{A} &= \mathbf{B}_1\mathbf{B}_1' + \cdots + \mathbf{B}_r\mathbf{B}_r' \\ \mathbf{M}'\mathbf{AM} &= \sum (\mathbf{M}'\mathbf{B}_i)(\mathbf{M}'\mathbf{B}_i)' = 0 \\ &\Rightarrow \mathbf{M}'\mathbf{B}_i = 0, \quad i = 1, \cdots, r\end{aligned}$$

となるので，(8b.2.2) から

$$\mathscr{U}'\mathbf{A}\mathscr{U} = \sum \mathbf{V}_i\mathbf{V}_i', \qquad \text{ただし，} E(\mathbf{V}_i) = \mathbf{M}'\mathbf{B}_i = 0$$

を得，定義によってこの分布はウィシャート分布で，かつ中心分布でるあことがわかる．

(ii) の結果をより直接的な形であらわせば，$\mathscr{U}'\mathbf{A}\mathscr{U} \sim W_p$ であるための必要十分条件は，$\mathbf{A}$ がベキ等であることであり，$\mathbf{AM} = 0$ のときその分布は中心分布となる．この結果を間接的な形で述べた理由は，そうすれば以下のような方法で1変量理論の既知の結果を一般化できるからである．

$\mathbf{Y}$ が独立な N_1 確率変数のベクトルで，2次形式 $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi^2$ であることが知られていると，$\mathbf{Y}$ に $\mathscr{U}\mathbf{L}$ を代入し，$\mathbf{L}$ についての2次形式 $\mathbf{L}'\mathscr{U}'\mathbf{A}\mathscr{U}\mathbf{L}$ を求めたうえで，$\mathbf{L}$ を除くことにより，そのなかの行列は $\mathscr{U}'\mathbf{A}\mathscr{U} \sim W_p$ であるといえる．

(ii) の議論から (iii) で述べる結果の証明は容易である．

(iii) (ii) と同じ条件と記号の下で，行列 $\mathcal{U}'\mathbf{A}_1\mathcal{U}$ と $\mathcal{U}'\mathbf{A}_2\mathcal{U}$ が独立にウィシャート分布に従うための必要十分条件は，$\mathbf{Y}'\mathbf{A}_1\mathbf{Y}$ と $\mathbf{Y}'\mathbf{A}_2\mathbf{Y}$ が任意の $\mathbf{L}$ にたいして独立な χ^2 分布に従うことである．さらに，$\mathbf{Y}'\mathbf{B}$ と $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ が任意の $\mathbf{L}$ にたいして独立に N_1 および χ^2 分布に従うならば，$\mathcal{U}'\mathbf{B}$ と $\mathcal{U}'\mathbf{A}\mathcal{U}$ は独立に N_p および W_p 分布に従う．

(iii) の結果は，$\mathcal{U}'\mathbf{A}_1\mathcal{U}$ と $\mathcal{U}'\mathbf{A}_2\mathcal{U}$ が独立に分布する（ウィシャート分布であってもなくても）ための必要十分条件は $\mathbf{A}_1\mathbf{A}_2=\mathbf{0}$ であり，$\mathcal{U}'\mathbf{B}$ と $\mathcal{U}'\mathbf{A}\mathcal{U}$ が独立に分布する（後者がウィシャート分布であってもなくても）ための必要十分条件は $\mathbf{B}'\mathbf{A}=\mathbf{0}$ である，という形式で述べることもできるであろう．

しかし，本書では，独立な χ^2 分布に従うことがわかっている互いに独立な N_1 確率変数の2次形式を考え，そこに含まれる行列を明確に規定することなしに，対応する結果を多変量の場合へ一般化したい．その直接的な例として (iv) の結果がある．われわれは本章を通じてこの方法を探索しつづけ，この主題に関して他の本で使われているような面倒な変数変換を用いて多変量の場合の分布を誘導するという直接的な試みをすべて避けたい．

(iv)　**標本平均と分散共分散行列の同時分布**

$\mathbf{U}_1,\cdots,\mathbf{U}_n$ は，互いに独立に同一の分布 $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ に従うとする．確率変数 $\mathbf{L}'\mathbf{U}_1,\cdots,\mathbf{L}'\mathbf{U}_n$ を考えると，これらは互いに独立に同一分布 $N_1(\mathbf{L}'\boldsymbol{\mu},\mathbf{L}'\boldsymbol{\Sigma}\mathbf{L})$ に従う．1変量理論から，任意の $\mathbf{L}$ にたいして標本平均の分布は

$$\frac{1}{n}\sum \mathbf{L}'\mathbf{U}_r=\mathbf{L}'\overline{\mathbf{U}}\sim N_1\left(\mathbf{L}'\boldsymbol{\mu},\frac{1}{n}\mathbf{L}'\boldsymbol{\Sigma}\mathbf{L}\right) \tag{8b.2.3}$$

となり，補正済平方和の分布は，

$$\sum_1^n(\mathbf{L}'\mathbf{U}_r)^2-n(\mathbf{L}'\overline{\mathbf{U}})^2=\mathbf{L}'(\sum \mathbf{U}_r\mathbf{U}_r'-n\overline{\mathbf{U}}\overline{\mathbf{U}}')\mathbf{L}$$

$$=\mathbf{L}'(S_{ij})\mathbf{L}\sim(\mathbf{L}'\boldsymbol{\Sigma}\mathbf{L})\chi^2(n-1),\ \text{中心}\ \chi^2\ \text{分布} \tag{8b.2.4}$$

となり，これらは互いに独立であることが知られている．したがって，$(S_{ij})=\sum \mathbf{U}_r\mathbf{U}_r'-n\overline{\mathbf{U}}\overline{\mathbf{U}}'$ を (8a.5.2) で定義した補正済平方和・積和とすれば，任意の $\mathbf{L}$ についての $\mathbf{L}'\overline{\mathbf{U}}$ と $\mathbf{L}'(S_{ij})\mathbf{L}$ の独立性から，($\mathbf{L}$ を落とすことにより) $\overline{\mathbf{U}}$ と (S_{ij}) は独立に分布することがわかる．(8b.2.3) から $\overline{\mathbf{U}}\sim N_p(\boldsymbol{\mu},(1/n)\boldsymbol{\Sigma})$ であり，(8b.2.4) から $(S_{ij})\sim W_p(n-1,\boldsymbol{\Sigma})$ すなわち中心ウィシャート分布に従う [(ii) を用いて]．

(v) $\mathbf{S}_1\sim W_p(k_1,\boldsymbol{\Sigma})$ と $\mathbf{S}_2\sim W_p(k_2,\boldsymbol{\Sigma})$ は互いに独立とすると，$\mathbf{S}=\mathbf{S}_1+\mathbf{S}_2\sim W_p(k_1+k_2,\boldsymbol{\Sigma})$ であり，非心ウィシャート分布にたいしても同様の結果が成り立つ．

互いに独立に同一分布 $N_p(0,\boldsymbol{\Sigma})$ に従う $\mathbf{U}_1,\cdots,\mathbf{U}_{k_1+k_2}$ を用いて，

$$\mathbf{S}_1=\sum_1^{k_1}\mathbf{U}_i\mathbf{U}_i'\quad\text{および}\quad\mathbf{S}_2=\sum_{k_1+1}^{k_1+k_2}\mathbf{U}_i\mathbf{U}_i'$$

と書くことができる．したがって，定義により

$$\mathbf{S}=\sum_1^{k_1+k_2}\mathbf{U}_i\mathbf{U}_i'\sim W_p(k_1+k_2,\boldsymbol{\Sigma}).$$

(vi) $\mathbf{S}\sim W_p(k,\boldsymbol{\Sigma})$ で $\mathbf{B}$ を $q\times p$ 行列とすれば，$\mathbf{BSB}'\sim W_q(k,\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')$．

$\mathbf{BU}_i\sim N_q(0,\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')$，$i=1,\cdots,k$ で，かつ互いに独立であるので，

$$\mathbf{S}=\sum_{i=1}^{k}\mathbf{U}_i\mathbf{U}_i'\Rightarrow\mathbf{BSB}'=\mathbf{B}(\sum\mathbf{U}_i\mathbf{U}_i')\mathbf{B}'$$

$$\mathbf{BSB}'=\sum_{i=1}^{k}(\mathbf{BU}_i)(\mathbf{BU}_i)'\sim W_q(k,\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}').$$

(vii) $W_p(\mathbf{S}\,|\,k,\boldsymbol{\Sigma})$ で $\mathbf{S}$ の密度関数をあらわし，$\mathbf{C}$ を p 次の非特異正方行列とすると，$\mathbf{S}_*=\mathbf{CSC}'$ の密度関数は

$$|\mathbf{C}|^{-p-1}W_p(\mathbf{C}^{-1}\mathbf{S}_*\mathbf{C}^{-1\prime}\,|\,k,\boldsymbol{\Sigma})$$

となる．

この結果は，変数変換と，ヤコビアン

$$\frac{D(\mathbf{S})}{D(\mathbf{S}_*)}=|\mathbf{C}|^{-p-1}$$

の計算から得られる．

(viii) (S^{ij}) が $\mathbf{S}=(S_{ij})$ の逆行列であり，かつ $(\sigma_{ij})=\boldsymbol{\Sigma}$ の逆行列 (σ^{ij}) が存在するとすれば，$\mathbf{S}\sim W_p(k,\boldsymbol{\Sigma})$ なら

(a) $\dfrac{\sigma^{pp}}{S^{pp}}\sim\chi^2(k-\overline{p-1})$ で，これは (S_{ij})，$i,j=1,\cdots,p-1$ と独立である．

(b) $\dfrac{\mathbf{L}'\boldsymbol{\Sigma}^{-1}\mathbf{L}}{\mathbf{L}'\mathbf{S}^{-1}\mathbf{L}}\sim\chi^2(k-\overline{p-1})$ が任意の定数ベクトル $\mathbf{L}$ にたいして成り立つ．

k 個の独立な確率変数 $\mathbf{U}_1,\cdots,\mathbf{U}_k$ がそれぞれ $N_p(0,\boldsymbol{\Sigma})$ に従って分布すると考える．$U_{1j},\cdots,U_{p-1,j}$ が与えられたときの U_{pj} の条件つき分布は

$$N_1(\beta_1U_{1j}+\cdots+\beta_{p-1}U_{p-1,j},1/\sigma^{pp})\quad j=1,\cdots,k \tag{8b.2.4}$$

である ((8a.2.16)，(8a.2.17) 参照)．(8b.2.4) の結果はガウス＝マルコフ模型での最小2乗法におけるものと同様である．したがって最小2乗法の第1基本定理 [(i), **3b.5**] を用いると，

$$R_0^2=\min_{\beta}\sum_1^k(U_{pj}-\beta_1U_{1j}-\cdots-\beta_{p-1}U_{p-1,j})^2 \tag{8b.2.5}$$

は，(U_{ij})，$i=1,\cdots,p-1$；$j=1,\cdots,k$ の階数が r のとき，$(\sigma^{pp})^{-1}\chi^2(k-r)$ に従って分布する．R_0^2 の分布は条件付きであるが，条件変数 (U_{ij})，$i=1,\cdots,p-1$；$j=1,\cdots,$

k を含まないので，その分布は同時に無条件であり，(U_{ij})，$i=1,\cdots,p-1$；$j=1,\cdots,k$ と独立である．

$\mathbf{S}\sim W_p(k,\Sigma)$ で，かつ $|\Sigma|\neq 0$ であるので，

$$\mathbf{S}=\sum_{j=1}^{k}\mathbf{U}_j\mathbf{U}_j',\qquad \mathbf{U}_i\sim N_p(\mathbf{0},\Sigma),\ \ j=1,\cdots,k \tag{8b.2.6}$$

の形に $\mathbf{S}$ をあらわすことができ，したがって，$k>(p-1)$ ならば，(U_{ij})，$i=1,\cdots,p-1$；$j=1,\cdots,k$ の階数は確率1で $p-1$ となる．上の結果 (a) は，$R_0^2=1/S^{pp}$ ならば成り立つが，これは1変量理論ですでに証明されている ((4a.5.2) 参照)．

U_{pi} のかわりに，$\mathbf{U}_i$ の任意の成分を選んで，その条件付き分布を考えることもできる．したがって σ^{ii}/S^{ii} の分布はすべての i にたいして同じである．

結果 (b) は，$\mathbf{B}$ が直交行列でその第1行が $\mathbf{L}$ に比例するとき，$\mathbf{BSB}'$ を考えるならば (a) から導かれる．(vi) の結果から，$\mathbf{BSB}'\sim W_p(k,\mathbf{B}\Sigma\mathbf{B}')$ であり，さらに

$$(\mathbf{BSB}')^{-1}=\mathbf{BS}^{-1}\mathbf{B}',\qquad (\mathbf{B}\Sigma\mathbf{B}')^{-1}=\mathbf{B}\Sigma^{-1}\mathbf{B}'$$

である．$\mathbf{BS}^{-1}\mathbf{B}'$ の第1対角要素は $\mathbf{L}'\mathbf{S}^{-1}\mathbf{L}$ に比例し，$\mathbf{B}\Sigma^{-1}\mathbf{B}'$ の第1要素も $\mathbf{L}'\Sigma^{-1}\mathbf{L}$ に比例し，同じ比例定数をもつ．したがって，結果 (a) を適用すると，

$$(\mathbf{L}'\Sigma^{-1}\mathbf{L}/\mathbf{L}'\mathbf{S}^{-1}\mathbf{L})\sim\chi^2(k-\overline{p-1})$$

である．

(ix)　$\mathbf{S}\sim W_p(k,\Sigma)$ とし，$\mathbf{S}$ の分割

$$\mathbf{S}=\begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix} \tag{8b.2.7}$$

を考えよう．ただし $\mathbf{S}_{11},\mathbf{S}_{12},\mathbf{S}_{22}$ は $r\times r$，$r\times s$，および $s\times s$ 行列で，$r+s=p$ である．すると，

$$\mathbf{S}_{22}-\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\sim W_s(k-r,\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

となる．ここで，Σ_{ij} は (8b.2.7) と同様に Σ の分割でつくられる．

$\mathbf{S}\sim W_p(k,\Sigma)$ であるから

$$\mathbf{S}=\sum_{1}^{k}\mathbf{U}_i\mathbf{U}_i',\qquad \mathbf{U}_i\sim N_p(\mathbf{0},\Sigma)$$

と書ける．$\mathbf{U}_i'=(\mathbf{U}_{1i}'\,\vdots\,\mathbf{U}_{2i}')$ の分割と，$r+1$ 次元の確率変数 $(\mathbf{U}_{1i}'\,\vdots\,\mathbf{L}'\mathbf{U}_{2i}')$ を考えよう．これは，ウィシャート行列

$$\begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\mathbf{L}\\ \mathbf{L}'\mathbf{S}_{21} & \mathbf{L}'\mathbf{S}_{22}\mathbf{L}\end{pmatrix}$$

をもつ．ただし $\mathbf{L}$ は s 個の要素の定数列ベクトルである．

(viii) の結果 (a) を用いると，

$$\begin{vmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\mathbf{L}\\ \mathbf{L}'\mathbf{S}_{21} & \mathbf{L}'\mathbf{S}_{22}\mathbf{L}\end{vmatrix}\div|\mathbf{S}_{11}|\sim c\chi^2(k-r)$$

であることがわかる．この左辺の式は

$$\mathbf{L}'\mathbf{S}_{22}\mathbf{L}-\mathbf{L}'\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{L}=\mathbf{L}'(\mathbf{S}_{22}-\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})\mathbf{L} \tag{8b.2.8}$$

と書け，同様に $c=\mathbf{L}'(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\mathbf{L}$ となる．(8b.2.8) から $\mathbf{L}$ を落とすと，その分布は次のとおりとなる．

$$\mathbf{S}_{22}-\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\sim W_s(k-r,\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).$$

(x)　$\mathbf{S}\sim W_p(k,\Sigma)$ で $|\Sigma|\neq 0$ とすると，$|\mathbf{S}|/|\Sigma|$ は，自由度 $k-p+1,\cdots,k-1,k$ の p 個の独立な中心 χ^2 確率変数の積の分布に従う．

次の分割（$|\mathbf{S}|_r=|(S_{ij})|$，$i,j=1,\cdots,r$ とあらわす）に注目する．

$$\frac{|\mathbf{S}|}{|\Sigma|}=\frac{|\mathbf{S}|_p}{|\mathbf{S}|_{p-1}}\cdot\frac{|\Sigma|_{p-1}}{|\Sigma|_p}\times\frac{|\mathbf{S}|_{p-1}}{|\mathbf{S}|_{p-2}}\cdot\frac{|\Sigma|_{p-2}}{|\Sigma|_{p-1}}\times\cdots\times\frac{|\mathbf{S}|_1}{|\Sigma|_1}$$

ここで，(viii) の (a) を適用すると，各因子はそれぞれの自由度の独立な χ^2 分布に従う．

(xi)　$\mathbf{S}_1\sim W_p(k_1,\Sigma)$ と $\mathbf{S}_2\sim W_p(k_2,\Sigma)$ は互いに独立であるとする．$k_1\geqslant p$ ならば，$\Lambda=|\mathbf{S}_1|/|\mathbf{S}_1+\mathbf{S}_2|$ は p 個の独立なベータ確率変数の積の分布に従い，その母数は

$$\left(\frac{k_1-p+1}{2},\ \frac{k_2}{2}\right),\ \left(\frac{k_1-p+2}{2},\ \frac{k_2}{2}\right),\ \ \cdots,\ \ \left(\frac{k_1}{2},\ \frac{k_2}{2}\right)$$

となる．$k_2=1$ という特殊な場合，ベータ確率変数の積は，次のパラメータ

$$\left(\frac{k_1-p+1}{2},\ \frac{p}{2}\right)$$

をもつ単一のベータ確率変数と同一の分布に従う（3a.3 の (iii) を用いる）．

互いに独立なベクトル確率変数 $\mathbf{U}_1,\cdots,\mathbf{U}_{k_1},\mathbf{U}_{k_1+1},\cdots,\mathbf{U}_{k_1+k_2}$ を考え，

$$\mathbf{S}_1=\sum_{1}^{k_1}\mathbf{U}_i\mathbf{U}_i'\quad\text{および}\quad\mathbf{S}_2=\sum_{k_1+1}^{k_1+k_2}\mathbf{U}_i\mathbf{U}_i'$$

とする．$U_{1i},\cdots,U_{p-1,i}$ を与えたときの U_{pi} の条件つき分布は

$$N_1\left(\beta_1U_{1i}+\cdots+\beta_{p-1}U_{p-1,i},\ \frac{1}{\sigma^{pp}}\right),\quad i=1,\cdots,k_1+k_2$$

である．さて，最小2乗法の第3基本定理（3b.5 参照）によると，

$$B_p=\frac{\min\limits_{\beta}\sum_{1}^{k_1}(U_{pi}-\beta_1U_{1i}-\cdots-\beta_{p-1}U_{p-1,i})^2}{\min\limits_{\beta}\sum_{1}^{k_1+k_2}(U_{pi}-\beta_1U_{1i}-\cdots-\beta_{p-1}U_{p-1,i})^2} \tag{8b.2.9}$$

は

$$B\left(\frac{k_1-p+1}{2},\frac{k_2}{2}\right)$$

に従って分布する．また，式 (8b. 2. 5) を使えば，(8b. 2. 9) の分子は $|\mathbf{S}_1|_p \div |\mathbf{S}_1|_{p-1}$，分母は $|\mathbf{S}_1+\mathbf{S}_2|_p \div |\mathbf{S}_1+\mathbf{S}_2|_{p-1}$ である．さらに Λ の分割

$$\frac{|\mathbf{S}_1|}{|\mathbf{S}_1+\mathbf{S}_2|}=\left[\frac{|\mathbf{S}_1|_p}{|\mathbf{S}_1|_{p-1}} \div \frac{|\mathbf{S}_1+\mathbf{S}_2|_p}{|\mathbf{S}_1+\mathbf{S}_2|_{p-1}}\right]\times\left[\frac{|\mathbf{S}_1|_{p-1}}{|\mathbf{S}_1|_{p-2}} \div \frac{|\mathbf{S}_1+\mathbf{S}_2|_{p-1}}{|\mathbf{S}_1+\mathbf{S}_2|_{p-2}}\right]$$

$$\times\cdots\times\left[\frac{|\mathbf{S}_1|_1}{|\mathbf{S}_1+\mathbf{S}_2|_1}\right]=B_p\times B_{p-1}\times\cdots\times B_1 \qquad (8b.2.10)$$

を用いる．ただし $B_p, B_{p-1}, \cdots, B_1$ はすべて互いに独立である．(8b. 2. 9) の分布は (U_{ij})，$i \neq p$ と独立であるので，B_p は残りの $B_i, i<p$ と独立である．同様にして B_{p-1} は B_i，$i<p-1$ と独立である．以下同様．B_j の分布の母数は (8b. 2. 9) の p を j にかえて

$$\left(\frac{k_1-j+1}{2}, \frac{k_2}{2}\right)$$

である．われわれは Λ 統計量の分布を $\Lambda(p, k_1, k_2)$ であらわす．

（xii）**ホテリングの分布**　$\mathbf{S}\sim W_p(k,\Sigma)$ と $\mathbf{d}\sim N_p(\boldsymbol{\delta}, c^{-1}\Sigma)$ は互いに独立であるとする．ホテリングの一般化 T^2 統計量は

$$T^2=ck\mathbf{d}'\mathbf{S}^{-1}\mathbf{d}=\frac{k\mathbf{d}'\mathbf{S}^{-1}\mathbf{d}}{\mathbf{d}'\Sigma^{-1}\mathbf{d}}c\mathbf{d}'\Sigma^{-1}\mathbf{d}$$

で定義される．与えられた $\mathbf{d}$ にたいして，(viii) の (b) を用いると，

$$\frac{\mathbf{d}'\Sigma^{-1}\mathbf{d}}{\mathbf{d}'\mathbf{S}^{-1}\mathbf{d}}\sim\chi^2(k-p+1) \qquad (8b.2.11)$$

である．分布 (8b. 2. 11) は $\mathbf{d}$ を含まないので，比 $\mathbf{d}'\Sigma^{-1}\mathbf{d}/\mathbf{d}'\mathbf{S}^{-1}\mathbf{d}$ は $\mathbf{d}$ と独立に分布する．$\mathbf{d}\sim N_p(\boldsymbol{\delta}, c^{-1}\Sigma)$ であるので

$$c\mathbf{d}'\Sigma^{-1}\mathbf{d}\sim\chi^2(p, c\tau^2), \qquad (8b.2.12)$$

ただし，$\tau^2=\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta}$ である．分布 (8b. 2. 11) と (8b. 2. 12) は互いに独立である．したがって，統計量 T^2/k は，非心カイ2乗 $\chi^2(p, c\tau^2)$ とこれと独立な中心カイ2乗 $\chi^2(k-p+1)$ の比である．したがって，

$$\frac{k-p+1}{p}\frac{T^2}{k}\sim F(p, k-p+1, c\tau^2) \qquad (8b.2.13)$$

となる．これは，第3章問題**17. 1**で定義した非心 F 分布である．

$\boldsymbol{\delta}=0$ ならば，2つの χ^2 はともに中心分布となり，

$$\frac{k-p+1}{p}\frac{T^2}{k}\sim F(p, k-p+1) \qquad (8b.2.14)$$

を得る．これは中心 F 分布 (3a. 2. 13) である．

互いに独立な χ^2 分布の比として T^2 をあらわし，その分布の誘導をエレガントに行うこのやり方は Wijsman(1957) に負うものである．また Bowker(1960) も参照せよ．初期の論文としては Bose and Roy(1938)，Hotelling(1931)，Hsu(1938, 1947)，K. S. Rao

(1951) および Rao(1946f, 1949b) を見よ．

また，簡単な計算で，

$$\left(1+\frac{T^2}{k}\right)^{-1}=\frac{|\mathbf{S}|}{|\mathbf{S}+c\mathbf{dd}'|} \qquad (8b.2.15)$$

であることがわかる．ここで，$\boldsymbol{\delta}=0$ のとき，$c\mathbf{dd}'\sim W_p(1,\Sigma)$ である．よって，ホテリングの T^2 は，単調な変換をすれば，(xi) で考察した統計量 $\Lambda=|\mathbf{S}_1|/|\mathbf{S}_1+\mathbf{S}_2|$ の $k_2=1$ とした特殊な場合と見ることができる．(8b. 2. 15) の関係と，T^2 の分布 (8b. 2. 14) を用いると

$$\Lambda=\frac{|\mathbf{S}|}{|\mathbf{S}+c\mathbf{dd}'|}\sim B\left(\frac{k-p+1}{2}, \frac{p}{2}\right)$$

が示される．これは (xi) とは独立に導かれたものである．したがって，(8b. 2. 15) の関係を用いて，T^2 の分布 (8b. 2. 14) を得ることもできるであろう．すなわち，われわれはホテリングの分布にたいして，2通りの簡潔な証明を得た．

（xiii）**固有方程式の根**　フィッシャーは，固有方程式

$$|\mathbf{S}_1-\lambda(\mathbf{S}_1+\mathbf{S}_2)|=0 \qquad (8b.2.16)$$

の根に基づくいくつかの仮説検定を導入した．ここで，$\mathbf{S}_1\sim W_p(k_1,\Sigma)$ と $\mathbf{S}_2\sim W_p(k_2,\Sigma)$ は互いに独立である．この根の正確な分布を導くことはしないが，その密度関数が p, k_1 および k_2 のみを含み，未知の Σ を含まないことはわかるであろう．密度関数の誘導については，Fisher(1939)，Girshick(1939)，Hsu(1939)，Roy(1939) の原論文，および James(1954) による興味ある証明を参照せよ．詳しい証明は Anderson(1958)，Roy(1957) および Wilks(1962) の著書にある．われわれの研究では密度関数は必要がない．Bagai, Constantine, Khatri, Krishnaih, Mathai, Pillai などによる，(8b. 2. 16) の根に関する研究については，Rao(1972a) を参照されたい．

8c　分散共分散分析

8c. 1　多重測定にたいするガウス＝マルコフ模型

第4章で述べたガウス＝マルコフ模型における最小2乗法は，1つの実験単位または個体について，1つの測定値が得られる（つまり，特定の性質の値が得られる）場合に適用可能である．その測定値（すなわち1次元確率変数）の期待値は，未知母数の1次関数で，それを構成する係数は採用された特定の実験計画によって定められると考えた．選ばれた同じ個体について，第2の測定値がとられる（つまり，もう1つの特性が測られる）場合

には，その期待値は，第1の測定値のそれと同じ1次関数であるが，未知母数の組は異なる，と考えるのが自然である．このように，母数は調べられる特性によって変わるが，その1次関数の係数は採用される実験計画によって定まるのである．

したがって，実験ブロック i で特定の処理 j を受けた植物の収量の期待値を，i ブロックと j 処理の効果をあらわす $\beta_{i1}+\tau_{j1}$ と書くなら，植物の草丈のような第2の特性の期待値は，類似の効果をあらわす他の2つの母数の和 $\beta_{i2}+\tau_{j2}$ と書けるであろう．

いま，p 種の特性が，n 個の独立な実験単位または個体のおのおので観察されるとしよう．その観測値はつぎのように書ける．

個　体	特　　性 (1)　(2) … (p)	
(1) ⋮ (n)	$\begin{pmatrix} U_{11} & U_{21} & \cdots & U_{p1} \\ \vdots & \vdots & \cdots & \vdots \\ U_{1n} & U_{2n} & \cdots & _{pn} \end{pmatrix} = \mathscr{U}$	(8c. 1. 1)

$\mathscr{U}$ の i 番目の列を $\mathbf{Y}_i$ であらわそう．すると，われわれの考える模型（モデル）は

$$E(\mathbf{Y}_i)=\mathbf{X}\boldsymbol{\beta}_i,\quad i=1,\cdots,p,$$

$$\mathrm{Cov}(\mathbf{Y}_i,\mathbf{Y}_j)=\sigma_{ij}\mathbf{I},\quad i,j=1,\cdots,p \tag{8c. 1. 2}$$

であり，いいかえると，そのモデルは各 $\mathbf{Y}_i$ についてガウス＝マルコフ $(\mathbf{Y}_i,\mathbf{X}\boldsymbol{\beta}_i,\sigma_{ii}\mathbf{I})$ で，未知ベクトル母数 $\boldsymbol{\beta}_i$ は i ごとに特有であるが，すべての i について共通な $n\times m$ の計画行列 $\mathbf{X}$ をもっている．i 番目の特性の分散は σ_{ii} である．調べられる特性間の従属性は

$$\mathrm{Cov}(\mathbf{Y}_i,\mathbf{Y}_j)=\sigma_{ij}\mathbf{I} \tag{8c. 1. 3}$$

という関係であらわされる．ただし，σ_{ij} は個体の i 番目と j 番目の特性間の共分散である．

$\mathbf{B}$ を $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_p$ を列ベクトルとする $m\times p$ 行列とし，$\boldsymbol{\Sigma}=(\sigma_{ij})$ とすると，(8c. 1. 2) と (8c. 1. 3) をもつ多変量模型は簡単に

$$(\mathscr{U},\mathbf{XB},\boldsymbol{\Sigma}\otimes\mathbf{I}) \tag{8c. 1. 4}$$

と表現できる．ただし，クロネッカー積行列 $\boldsymbol{\Sigma}\otimes\mathbf{I}$ は，$\mathscr{U}$ の列を次々に下に書き並べて得られる np 次元のベクトル変数 $\mathbf{Y}$，すなわち $\mathbf{Y}'=(\mathbf{Y}_1'\,\vdots\cdots\vdots\,\mathbf{Y}_p')$ の共分散行列と解釈される．

(8c. 1. 4) のモデルを，$\mathbf{Y}$，その期待値，および分散共分散行列を含む1変量ガウス＝マルコフ模型としてあらわすこともできる．$\boldsymbol{\beta}$ を $\boldsymbol{\beta}'=(\boldsymbol{\beta}_1'\,\vdots\,\boldsymbol{\beta}_2'\,\vdots\cdots\vdots\,\boldsymbol{\beta}_p')$ で定義される mp 次元ベクトルとする．このとき

$$E(\mathbf{Y})=(\mathbf{I}\otimes\mathbf{X})\boldsymbol{\beta},\qquad D(\mathbf{Y})=\boldsymbol{\Sigma}\otimes\mathbf{I} \tag{8c. 1. 5}$$

であるから，多変量線形模型は，第4章のガウス＝マルコフ模型の3つの要素

$$(\mathbf{Y},(\mathbf{I}\otimes\mathbf{X})\boldsymbol{\beta},\boldsymbol{\Sigma}\otimes\mathbf{I}) \tag{8c. 1. 6}$$

であらわすことができることがすぐわかる．そうすれば，第4章で述べた線形模型における推定についての諸結果を用いることもできる．しかしながら，この共分散の構造は特殊であるから，そのモデルは，(8c. 1. 4) または (8c. 1. 2) の形で，つまり，個々の特性についてはガウス＝マルコフで，特性間の相関を示すためには (8c. 1. 3) を追加して，あらわすほうがよい．この表現は，1変量理論と多変量理論の間の親密な関係を示すのに，大きな利点をもっている (8c. 2—8c. 5 参照)．

8c. 2　母 数 の 推 定

1変量のガウス＝マルコフ模型で行ったように，変数の1次と2次の積率のもつ構造 (8c. 1. 2) と (8c. 1. 3) だけを用い，特定の分布を仮定しないで未知母数の推定を考えよう．

まず，i 番目の特性に固有の母数のみを含む $\mathbf{P}'\boldsymbol{\beta}_i$ という型の母数関数を，すべての観測値の1次関数によって，その推定量が (a) 不偏で，(b) 最小分散をもつように推定する問題を考える．すべての観測値の1次関数は

$$\mathbf{A}_1'\mathbf{Y}_1+\cdots+\mathbf{A}_p'\mathbf{Y}_p \tag{8c. 2. 1}$$

と書ける．不偏性の基準 (a) を用いると，

$$E(\mathbf{A}_1'\mathbf{Y}_1)+\cdots+E(\mathbf{A}_p'\mathbf{Y}_p)=\mathbf{A}_1'\mathbf{X}\boldsymbol{\beta}_1+\cdots+\mathbf{A}_p'\mathbf{X}\boldsymbol{\beta}_p=\mathbf{P}'\boldsymbol{\beta}_i$$

$$\Rightarrow \mathbf{A}_i'\mathbf{X}=\mathbf{P}',\quad \mathbf{A}_j'\mathbf{X}=0,\quad j\neq i \tag{8c. 2. 2}$$

であることがわかる．$\mathbf{A}_i'\mathbf{X}=\mathbf{P}'$ であるから，$\mathbf{P}=\mathbf{X}'\mathbf{X}\boldsymbol{\lambda}$ をみたすようなベクトル $\boldsymbol{\lambda}$ が存在する．それを用いて1次関数 $\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}_i$ を考えよう［これは $E(\mathbf{Y}_i)=\mathbf{X}\boldsymbol{\beta}_i$ という前提の下で，$\mathbf{Y}_i$ だけを用いる $\mathbf{P}'\boldsymbol{\beta}_i$ の最小2乗推定量であることを知っている ((4a. 2. 3) 参照)］．この推定量は

$$E(\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}_i)=\boldsymbol{\lambda}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}_i=\mathbf{P}'\boldsymbol{\beta}_i \tag{8c. 2. 3}$$

であるから，不偏である．さらに

$$V(\textstyle\sum\mathbf{A}_j'\mathbf{Y}_j)=V(\sum\mathbf{A}_j'\mathbf{Y}_j-\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}_i)+V(\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}_i)$$

$$+2\,\mathrm{cov}(\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}_i,\textstyle\sum\mathbf{A}_j'\mathbf{Y}_j-\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}_i) \tag{8c. 2. 4}$$

である．しかし，(8c. 2. 2) を用いるならば，

$$\begin{aligned}\mathrm{cov}(\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}_i,\textstyle\sum\mathbf{A}_j'\mathbf{Y}_j-\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}_i)&=\sum_j\boldsymbol{\lambda}'\mathbf{X}'\mathrm{cov}(\mathbf{Y}_i,\mathbf{Y}_j)\mathbf{A}_j-\boldsymbol{\lambda}'\mathbf{X}'\mathrm{cov}(\mathbf{Y}_i,\mathbf{Y}_i)\mathbf{X}\boldsymbol{\lambda}\\&=\textstyle\sum\boldsymbol{\lambda}'\mathbf{X}'\mathbf{A}_j\sigma_{ij}-\boldsymbol{\lambda}'\mathbf{X}'\mathbf{X}\boldsymbol{\lambda}\sigma_{ii}\\&=\boldsymbol{\lambda}'\mathbf{X}'\mathbf{A}_i\sigma_{ii}-\boldsymbol{\lambda}'\mathbf{X}'\mathbf{X}\boldsymbol{\lambda}\sigma_{ii}=(\boldsymbol{\lambda}'\mathbf{P}-\boldsymbol{\lambda}'\mathbf{P})\sigma_{ii}=0.\end{aligned}$$

したがって，(8c. 2. 4) から

$$V(\textstyle\sum\mathbf{A}_j'\mathbf{Y}_j)\geqslant V(\boldsymbol{\lambda}'\mathbf{X}'\mathbf{Y}_i)$$

であることがわかる．すなわち，$\lambda'\mathbf{X}'\mathbf{Y}_i$ は $\mathbf{P}'\boldsymbol{\beta}_i$ の最小分散不偏推定量 m. v. u. e. (minimum variance unbiased estimator) である．それゆえ，$\mathbf{P}'\boldsymbol{\beta}_i$ の望ましい推定量は i 番目の特性だけについての観測値 $\mathbf{Y}_i$ に基づく最小2乗推定量と同じになる．いま，($\mathbf{Y}_i$ と $\boldsymbol{\beta}_i$ のみを考えるときの) 正規方程式

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta}_i = \mathbf{X}'\mathbf{Y}_i \tag{8c.2.5}$$

の解を $\hat{\boldsymbol{\beta}}_i$ とすると，$\mathbf{P}'\boldsymbol{\beta}_i$ の最小2乗推定量は，$\mathbf{P}'\hat{\boldsymbol{\beta}}_i$ と書ける．$\mathbf{P}'\boldsymbol{\beta}_i$ と $\mathbf{Q}'\boldsymbol{\beta}_i$ という2つの母数関数の推定量の分散と共分散は，最小2乗法のときと同じである．したがって，推定量 $\mathbf{P}'\hat{\boldsymbol{\beta}}_i$ と $\mathbf{Q}'\hat{\boldsymbol{\beta}}_i$ の分散と共分散は

$$V(\mathbf{P}'\hat{\boldsymbol{\beta}}_i) = a\sigma_{ii}, \qquad V(\mathbf{Q}'\hat{\boldsymbol{\beta}}_i) = b\sigma_{ii}$$
$$\mathrm{cov}(\mathbf{P}'\hat{\boldsymbol{\beta}}_i, \mathbf{Q}'\hat{\boldsymbol{\beta}}_i) = c\sigma_{ii} \tag{8c.2.6}$$

の形に書ける．ただし，係数 a, b, c は i と独立で，$\mathbf{X}'\mathbf{X}, \mathbf{P}, \mathbf{Q}$ だけに依存する (4a.3 参照)．たとえば，$(\mathbf{X}'\mathbf{X})^-$ を $\mathbf{X}'\mathbf{X}$ の一般逆行列とすると，$a = \mathbf{P}'(\mathbf{X}'\mathbf{X})^-\mathbf{P}$，$b = \mathbf{Q}'(\mathbf{X}'\mathbf{X})^-\mathbf{Q}$，および $c = \mathbf{P}'(\mathbf{X}'\mathbf{X})^-\mathbf{Q}$ となる．

いま，正規方程式 (8c.2.5) で $i = 1, \cdots, p$ とおいたときの解を $\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_p$ とし，$\mathbf{P}_i \in \mathscr{M}(\mathbf{X}')$，$i = 1, \cdots, p$ とすると，m. v. 推定量の和はまた m. v. 推定量である [(i), 5a.2 参照] ことから，

$$\mathbf{P}_1'\boldsymbol{\beta}_1 + \cdots + \mathbf{P}_p'\boldsymbol{\beta}_p \tag{8c.2.7}$$

は，

$$\mathbf{P}_1'\hat{\boldsymbol{\beta}}_1 + \cdots + \mathbf{P}_p'\hat{\boldsymbol{\beta}}_p$$

の m. v. u. e. であることがわかる．(8c.2.6) 式の構造から，

$$\mathrm{cov}(\mathbf{P}'\hat{\boldsymbol{\beta}}_i,\ \mathbf{P}'\hat{\boldsymbol{\beta}}_j) = a\sigma_{ij}$$
$$\mathrm{cov}(\mathbf{Q}'\hat{\boldsymbol{\beta}}_i,\ \mathbf{Q}'\hat{\boldsymbol{\beta}}_j) = b\sigma_{ij}$$
$$\mathrm{cov}(\mathbf{P}'\hat{\boldsymbol{\beta}}_i,\ \mathbf{Q}'\hat{\boldsymbol{\beta}}_j) = c\sigma_{ij}$$

であることがすぐわかる．ただし，a, b, c は (8c.2.6) のそれらと同じで，$\mathbf{P}$ と $\mathbf{Q}$ だけに依存するから，これらを用いて (8c.2.7) の分散を書きおろすことができる．そのためには，任意の i について，$\mathbf{P}_1'\hat{\boldsymbol{\beta}}_i, \cdots, \mathbf{P}_p'\hat{\boldsymbol{\beta}}_i$ の分散と共分散の式だけがわかればよいことになり，それは1変量の最小2乗法によって求められる．以上のことから，(8c.2.7) の分散は $\mathrm{tr}\,\mathbf{P}'(\mathbf{X}'\mathbf{X})^-\mathbf{P}\boldsymbol{\Sigma}$，ただし $\mathbf{P} = (\mathbf{P}_1 \vdots \cdots \vdots \mathbf{P}_p)$ であることがわかる．

σ_{ii} と σ_{ij} の推定量を求めるにはなんの困難もない．1変量の最小2乗法の理論から，σ_{ii} の不偏推定量は $R_0(i,i)/(n-r)$ であることを知っている．ただし，$r = R(\mathbf{X})$ で，

$$R_0(i,i) = \min_{\boldsymbol{\beta}_i}(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}_i)'(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}_i)$$
$$= (\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i)'(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i) = \mathbf{Y}_i'\mathbf{Y}_i - \mathbf{Y}_i'\mathbf{X}\hat{\boldsymbol{\beta}}_i. \tag{8c.2.8}$$

ここで，ベクトル確率変数 $\mathbf{Y}_i + \mathbf{Y}_j = \mathbf{Y}$ を考えよう．その期待値は $\mathbf{X}(\boldsymbol{\beta}_i + \boldsymbol{\beta}_j) = \mathbf{X}\boldsymbol{\beta}$

で，$D(\mathbf{Y}) = \sigma^2\mathbf{I}$，ただし $\sigma^2 = \sigma_{ii} + 2\sigma_{ij} + \sigma_{jj}$ である．$\boldsymbol{\beta}$ にたいする正規方程式は

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$$

で，解 $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_i + \hat{\boldsymbol{\beta}}_j$ の $\hat{\boldsymbol{\beta}}_i$ と $\hat{\boldsymbol{\beta}}_j$ は (8c.2.5) の i および j にたいする解である．σ^2 の不偏推定量は $R_0^2/(n-r)$ である．ただし，

$$R_0^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_j)'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_j)$$
$$= (\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i)'(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i) + (\mathbf{Y}_j - \mathbf{X}\hat{\boldsymbol{\beta}}_j)'(\mathbf{Y}_j - \mathbf{X}\hat{\boldsymbol{\beta}}_j) + 2(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i)'(\mathbf{Y}_j - \mathbf{X}\hat{\boldsymbol{\beta}}_j)$$
$$= R_0(i,i) \qquad + R_0(j,j) \qquad + 2R_0(i,j)$$

ただし，$R_0(i,j) = (\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i)'(\mathbf{Y}_j - \mathbf{X}\hat{\boldsymbol{\beta}}_j)$ とおいた．さて，

$$E(R_0^2) = (n-r)\sigma^2 = (n-r)(\sigma_{ii} + 2\sigma_{ij} + \sigma_{jj}) \quad \text{（1変量理論より）}$$
$$= E[R_0(i,i)] + E[R_0(j,j)] + 2E[R_0(i,j)]$$
$$= (\sigma_{ii} + \sigma_{jj})(n-r) + 2E[R_0(i,j)]$$
$$\Rightarrow E[R_0(i,j) \div (n-r)] = \sigma_{ij}. \tag{8c.2.9}$$

これより，$\boldsymbol{\Sigma}$ のすべての要素の不偏推定量を得たことになる．その式

$$R_0(i,j) = (\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i)'(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_j)$$
$$= \mathbf{Y}_i'\mathbf{Y}_j - \mathbf{Y}_i'\mathbf{X}\hat{\boldsymbol{\beta}}_j = \mathbf{Y}_i'\mathbf{Y}_j - \mathbf{Y}_j'\mathbf{X}\hat{\boldsymbol{\beta}}_i \tag{8c.2.10}$$

は，残差積和とよばれる．ちょうど，$R_0(i,i)$ を残差平方和とよぶのと同じである．その行列全体

$$\mathbf{R}_0 = \begin{pmatrix} R_0(1,1) & \cdots & R_0(1,p) \\ \cdots & \cdots & \cdots \\ R_0(p,1) & \cdots & R_0(p,p) \end{pmatrix} \tag{8c.2.11}$$

は，自由度 $n-r$ をもつ残差平方和積和行列 (residual sum of squares and products matrix) (残差 S. S. P. または残差 S. P. 行列) と呼ばれる．それは，1変量最小2乗法における残差平方和の一般化である．

正規方程式 $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}_j = \mathbf{X}'\mathbf{Y}_j$ の解は，$\mathbf{C}$ を $\mathbf{X}'\mathbf{X}$ の一般逆行列とすると，$\boldsymbol{\beta}_j = \mathbf{C}\mathbf{X}'\mathbf{Y}_j$ と書ける．すると，$R_0(i,j) = \mathbf{Y}_i'\mathbf{Y}_j - \mathbf{Y}_i'\mathbf{X}\mathbf{C}\mathbf{X}'\mathbf{Y}_j = \mathbf{Y}_i'(\mathbf{I} - \mathbf{X}\mathbf{C}\mathbf{X}')\mathbf{Y}_j$ となる．したがって S. P. 行列 $\mathbf{R}_0$ は，(8c.1.1) における観測値行列 $\mathscr{U}'$ と計画行列 $\mathbf{X}$ を用いて，陽に，

$$\mathbf{R}_0 = \mathscr{U}'(\mathbf{I} - \mathbf{X}\mathbf{C}\mathbf{X}')\mathscr{U} = \mathscr{U}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}']\mathscr{U}$$

と書ける．$R(\mathbf{X}) = m$ なら，$(\mathbf{X}'\mathbf{X})^- = (\mathbf{X}'\mathbf{X})^{-1}$，すなわち真の逆行列となる．

このようにして，モデル ((8c.1.2), (8c.1.3)) の母数の推定は，そこに含まれる確率変数に特別な分布を仮定せずに完結する．しかし，線形仮説を検定するためには，各 $\mathbf{U}_i$ について p 変量正規分布を仮定することになろう．

8c.3　線形仮説の検定，分散共分散分析

線形仮説の組，

$$\mathbf{H}'\boldsymbol{\beta}_i = \boldsymbol{\xi}_i, \qquad i = 1, \cdots, p \tag{8c. 3. 1}$$

を同時に検定する問題を考えよう．ここに，行列 $\mathbf{H}'$ は次数 $q \times m$，階数 q で，ベクトル $\boldsymbol{\xi}_i$ は所与とする．

この問題を，仮説 $\mathbf{H}'\boldsymbol{\beta} = \boldsymbol{\xi}$，ただし $\boldsymbol{\beta} = L_1\boldsymbol{\beta}_1 + \cdots + L_p\boldsymbol{\beta}_p$，$\boldsymbol{\xi} = L_1\boldsymbol{\xi}_1 + \cdots + L_p\boldsymbol{\xi}_p$ および確率変数のベクトル $\mathbf{Y} = L_1\mathbf{Y}_1 + \cdots + L_p\mathbf{Y}_p$ を考えることによって，1変量の問題に帰着させる．すると，1変量のガウス＝マルコフ模型，$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$，$D(\mathbf{Y}) = \sigma_L^2\mathbf{I}$ の下で，仮説 $\mathbf{H}'\boldsymbol{\beta} = \boldsymbol{\xi}$ を検定することになる．それに適切な検定基準は，平方和の2つの最小値，

$$R_0^2 = \min_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

$$R_1^2 = \min_{\mathbf{H}'\boldsymbol{\beta} = \boldsymbol{\xi}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{8c. 3. 2}$$

に基づくものである．帰無仮説 $\mathbf{H}'\boldsymbol{\beta} = \boldsymbol{\xi}$ の下では，$R_1^2 - R_0^2$ と R_0^2 は互いに独立に，それぞれ自由度 q と $n-r$ をもつ中心 χ^2 分布に従う，ただし $r = R(\mathbf{X})$.

$(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}_i)'(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}_i)$ を最小にする $\boldsymbol{\beta}_i$ の値を $\hat{\boldsymbol{\beta}}_i$ とし，前に定義したように，

$$\begin{aligned} R_0(i, j) &= (\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i)'(\mathbf{Y}_j - \mathbf{X}\hat{\boldsymbol{\beta}}_j) \\ &= \mathbf{Y}_i'\mathbf{Y}_j - \mathbf{Y}_i'\mathbf{X}\hat{\boldsymbol{\beta}}_j = \mathbf{Y}_i'\mathbf{Y}_j - \mathbf{Y}_j'\mathbf{X}\hat{\boldsymbol{\beta}}_i \end{aligned} \tag{8c. 3. 3}$$

と書く．容易にわかるように $\hat{\boldsymbol{\beta}} = L_1\hat{\boldsymbol{\beta}}_1 + \cdots + L_p\hat{\boldsymbol{\beta}}_p$ は $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ を最小にし，

$$\begin{aligned} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= [\textstyle\sum L_i(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i)]'[\textstyle\sum L_i(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i)] \\ &= \textstyle\sum\sum L_iL_j(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_i)'(\mathbf{Y}_j - \mathbf{X}\hat{\boldsymbol{\beta}}_j) \\ &= \textstyle\sum\sum L_iL_jR_0(i, j) = \mathbf{L}'\mathbf{R}_0\mathbf{L} \end{aligned} \tag{8c. 3. 4}$$

を得る．ただし $\mathbf{R}_0 = (R_0(i, j))$ は (8c. 2. 11) で定義した残差 S. P. 行列である．

同様に，仮説 $\mathbf{H}'\boldsymbol{\beta}_i = \boldsymbol{\xi}_i$ のもとで，$(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}_i)'(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}_i)$ を最小にする $\boldsymbol{\beta}_i$ の値を $\boldsymbol{\beta}_i^*$ とし，

$$R_1(i, j) = (\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}_i^*)'(\mathbf{Y}_j - \mathbf{X}\boldsymbol{\beta}_j^*), \qquad \mathbf{R}_1 = (R_1(i, j))$$

とおく．このときは，(8c. 3. 4) と同様に

$$R_1^2 = \min_{\mathbf{H}'\boldsymbol{\beta} = \boldsymbol{\xi}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{L}'\mathbf{R}_1\mathbf{L} \tag{8c. 3. 5}$$

となり，これから

$$R_1^2 - R_0^2 = \mathbf{L}'(\mathbf{R}_1 - \mathbf{R}_0)\mathbf{L}$$

を得る．行列 $\mathbf{R}_1 - \mathbf{R}_0$ は，仮説からの偏差による S. P. 行列と呼ばれる．こうして，1変量のときの残差平方和および仮説からの偏差の平方和と同様に，“残差”および“仮説からの偏差”をあらわす S. P. 行列 $\mathbf{R}_0$ および $\mathbf{R}_1 - \mathbf{R}_0$ という2つの行列を得る．このような行列の分解 $\mathbf{R}_1 = \mathbf{R}_0 + (\mathbf{R}_1 - \mathbf{R}_0)$ は，分散分析における，$R_1^2 = R_0^2 + (R_1^2 - R_0^2)$ の一般化として得られ，分散共分散分析 A. D. (Analysis of Dispersion) と呼ばれる．

$\mathbf{R}_0$ と $\mathbf{R}_1 - \mathbf{R}_0$ の同時分布　帰無仮説 (8c. 3. 1) が真なら，いかなる $\mathbf{L}$ にたいしても［R_0^2 と R_1^2 は (8c. 3. 2) の定義どおりとして]，〔$\sigma_L^2 = \mathbf{L}'\boldsymbol{\Sigma}\mathbf{L}$〕

$$R_0^2 \sim \sigma_L^2\chi^2(n-r), \qquad R_1^2 - R_0^2 \sim \sigma_L^2\chi^2(q)$$

となり，かつ両者は独立である．一方 $R_0^2 = \mathbf{L}'\mathbf{R}_0\mathbf{L}$，$R_1^2 = \mathbf{L}'\mathbf{R}_1\mathbf{L}$ である．これより，[(iii)，8b. 2] の結果を用いて ($\mathbf{L}$ を落とすと)，行列 $\mathbf{R}_0$ と $\mathbf{R}_1 - \mathbf{R}_0$ は互いに独立なウィシャート分布

$$\mathbf{R}_0 \sim W_p(n-r, \Sigma), \qquad \mathbf{R}_1 - \mathbf{R}_0 \sim W_p(q, \Sigma) \tag{8c. 3. 6}$$

に従うことがわかる．

帰無仮説 (8c. 3. 1) が真でないならば，各 L について，

$$R_0^2 \sim \sigma_L^2\chi^2(n-r), \qquad R_1^2 - R_0^2 \sim \sigma_L^2\chi^2(q, \cdot) \quad \text{(中心または非心)}$$

となり，互いに独立である (再び1変量理論を用いる)．これより，行列 $\mathbf{R}_0$ と $\mathbf{R}_1 - \mathbf{R}_0$ は互いに独立なウィシャート分布

$$\mathbf{R}_0 \sim W_p(n-r, \Sigma), \qquad \mathbf{R}_1 - \mathbf{R}_0 \sim W_p(q, \Sigma, \cdot) \quad \text{(非心)}$$

に従うことがわかる．後者は $R_1^2 - R_0^2$ がいかなる $\mathbf{L}$ についても中心分布に従うときにのみ中心分布をする．したがって，帰無仮説 (8c. 3. 1) からの偏差は，行列 $\mathbf{R}_0$ と $\mathbf{R}_1 - \mathbf{R}_0$ を比較することによっても検出できる．

検定基準　帰無仮説 (8c. 3. 1) が真ならば，どの $\mathbf{L}$ にたいしても，仮説 $\mathbf{H}'\boldsymbol{\beta} = \boldsymbol{\xi}$ は真である．適切な検定は最小2乗法に基づく分散分析 F 検定

$$F = \frac{R_1^2 - R_0^2}{q} \div \frac{R_0^2}{n-r}$$

となる．ただし，R_0^2, R_1^2 は (8c. 3. 2) で定義したものである．

F のかわりに，統計量

$$B = \frac{R_0^2}{R_0^2 + (R_1^2 - R_0^2)} = \frac{R_0^2}{R_1^2} = \frac{1}{1 + \dfrac{qF}{n-r}}$$

を用いることもできる．この B は，F の単調減少関数で，次のベータ分布をする．

$$B\left(\frac{q}{2}, \frac{n-r}{2}\right)$$

B の値が小さくなると，有意性を示すことになる．$\mathbf{L}$ であらわすと，

$$B = \frac{\mathbf{L}'\mathbf{R}_0\mathbf{L}}{\mathbf{L}'\mathbf{R}_1\mathbf{L}} \tag{8c. 3. 7}$$

となる．これまでは $\mathbf{L}$ は任意であるが固定されていた．

仮説 (8c. 3. 1) からの偏差の効果は，B の値を (確率的に) 小さくする方向に現れることがすでにわかった．そこで，仮説を疑おうとするときには，B が最小の値をとるように $\mathbf{L}$ を選ぶことにしよう．そのとき，検定基準は，

$$\lambda = \min_{\mathbf{L}} B = \min_{\mathbf{L}} \frac{\mathbf{L}'\mathbf{R}_0\mathbf{L}}{\mathbf{L}'\mathbf{R}_1\mathbf{L}}$$

となる．つまり，λ は固有方程式

$$|\mathbf{R}_0 - \lambda\mathbf{R}_1| = 0 \tag{8c. 3. 8}$$

の最小根となる．

帰無仮説（8c. 3. 1）からの偏差が，方程式（8c. 3. 8）の根にどの程度反映するかを評価するために，$\mathbf{R}_0$ と $\mathbf{R}_1$ をその期待値におきかえよう．ここでも1変量の公式を使って，その期待値を求めることができる．（4b. 2. 3）で，R_0^2 と $R_1^2 - R_0^2$ の期待値は，

$$E(R_0^2) = (n-r)\sigma_L^2 = E(\mathbf{L}'\mathbf{R}_0\mathbf{L})$$

$$E(R_1^2 - R_0^2) = q\sigma_L^2 + (\mathbf{H}'\boldsymbol{\beta} - \boldsymbol{\xi})'\mathbf{D}^{-1}(\mathbf{H}'\boldsymbol{\beta} - \boldsymbol{\xi}) = E[\mathbf{L}'(\mathbf{R}_1 - \mathbf{R}_0)\mathbf{L}]$$

で，この $\mathbf{D}$ は $\mathbf{H}'\boldsymbol{\beta} - \boldsymbol{\xi}$ の推定量の分散共分散行列であることを知った．σ_L^2 と $\mathbf{H}'\boldsymbol{\beta} - \boldsymbol{\xi}$ を $\mathbf{L}$ であらわすと，

$$E(\mathbf{L}'\mathbf{R}_0\mathbf{L}) = (n-r)\mathbf{L}'\boldsymbol{\Sigma}\mathbf{L} \tag{8c. 3. 9}$$

$$\begin{aligned} E[\mathbf{L}'(\mathbf{R}_1 - \mathbf{R}_0)\mathbf{L}] &= q\mathbf{L}'\boldsymbol{\Sigma}\mathbf{L} + \left[\sum_1^p L_i(\mathbf{H}'\boldsymbol{\beta}_i - \xi_i)\right]'\mathbf{D}^{-1}\left[\sum_1^p L_i(\mathbf{H}'\boldsymbol{\beta}_i - \xi_i)\right] \\ &= q\mathbf{L}'\boldsymbol{\Sigma}\mathbf{L} + \mathbf{L}'\boldsymbol{\Delta}\mathbf{L} \end{aligned} \tag{8c. 3. 10}$$

となる．ただし，帰無仮説（8c. 3. 1）が真のとき $\boldsymbol{\Delta} = \mathbf{0}$，そうでなければ $\boldsymbol{\Delta}$ は非負定符号である．（8c. 3. 9）と（8c. 3. 10）はいかなる $\mathbf{L}$ についても真であるから，

$$E(\mathbf{R}_0) = (n-r)\boldsymbol{\Sigma}$$

$$E(\mathbf{R}_1 - \mathbf{R}_0) = q\boldsymbol{\Sigma} + \boldsymbol{\Delta}, \qquad E(\mathbf{R}_1) = \boldsymbol{\Delta} + (n - r + q)\boldsymbol{\Sigma}$$

となる．方程式（8c. 3. 8）で $\mathbf{R}_0$ と $\mathbf{R}_1$ をその期待値でおきかえると，

$$|(n-r)\boldsymbol{\Sigma} - \lambda^*(\boldsymbol{\Delta} + \overline{n - r + q}\,\boldsymbol{\Sigma})| = 0$$

と書け，それはさらに別の形

$$|\boldsymbol{\Delta} - \theta^*\boldsymbol{\Sigma}| = 0, \qquad \theta^* = \frac{(n-r) - \lambda^*(n-r+q)}{\lambda^*}$$

に書ける．

λ^* の小さい値は θ^* の大きい値に対応する．零でない根 θ_i^* の数は $\boldsymbol{\Delta}$ の階数に依存し，したがって，$\boldsymbol{\Delta} = \mathbf{0}$ からのずれは，零でないすべての根 θ_i^* またはそれに対応する λ_i^* の値に反映する．もし $\boldsymbol{\Delta} = \mathbf{0}$ なら，すべての θ_i^* は零となり，また，$\boldsymbol{\Delta}$ の階数が1のときは，偏差の全体は，最小の λ^* に対応するただ1つの非零根 θ^* に集中する．しかし，対立仮説 $\boldsymbol{\Delta}$ について何も知られていないときは，帰無仮説からのずれを検出するのに，2つ以上の根を用いるのが有益であろう．方程式（8c. 3. 8）の根からは，根 λ^* の推定値が得られるにすぎない．しかし，偏差行列に関する真の根の研究により，$\boldsymbol{\Delta}$ に関する仮説の検定において，推定根の使用が妥当性をもつことが示されている．

（8c. 3. 8）の最小根は，帰無仮説からの最大の偏差を表示する単一の根ではあるが，"残差" 行列 $\mathbf{R}_0$ と "仮説からの偏差" の行列 $\mathbf{R}_1 - \mathbf{R}_0$ との適切な比較を与えないこともある．それゆえ，すべての根のある（対称）関数で，関係する行列の全体としての比較を与えるようなものを考える．しかしながら，個々の根も，$\boldsymbol{\Delta}$ の階数を調べるのに，すでに示したように，重要な役割を果たす．そのような検定のくわしい議論は **8c. 6** に示す．

すべての根を含む検定基準として考えられる統計量は，

（a）$\Lambda = \lambda_1\lambda_2\cdots\lambda_p = |\mathbf{R}_0| \div |\mathbf{R}_1|$，　すべての根の積，

（b）$\dfrac{1-\lambda_1}{\lambda_1} + \cdots + \dfrac{1-\lambda_p}{\lambda_p}$,

（c）$\dfrac{1-\lambda_1}{\lambda_1} \times \dfrac{1-\lambda_2}{\lambda_2} \times \cdots \times \dfrac{1-\lambda_p}{\lambda_p}$,

などである．第1の基準は，Wilks(1932) によって，ある特殊の場合に尤度比の原理を適用して導かれ，後に Bartlett(1947) によって，多変量解析に広く用いられるように拡張された Λ 基準として有名である．

$\mathbf{R}_0$ と $\mathbf{R}_1$ の計算には，1変量分散分析（A. V.）で用いられる式以外の新しい公式は何も用いられていないことに注意せよ．A. V. 表に書き込まれる種々の値を求めるための公式は多くの場合に利用できるから，そのような場合の分散共分散分析（A. D.）への拡張は容易である．p 種の測定値 $\mathbf{U}$ の1次関数 $\mathbf{L}'\mathbf{U}$ を考え，合成係数 $\mathbf{L}$ の2次形式である R_0^2 と R_1^2 を求めるために，A. V. の既知の公式を適用するだけでよい．$\mathbf{L}$ のこれらの2次形式の行列は，正確に $\mathbf{R}_0$ と $\mathbf{R}_1$ である．帰無仮説の下では，$\Lambda = |\mathbf{R}_0| \div |\mathbf{R}_1|$ の分布は［(xi), **8b. 2**］で示したように，$\Lambda(p, n-r, q)$ である．

8c. 4　付加情報にたいする検定

8c. 3 では，各特性に関係する母数の同時仮説を考えた．たとえば，ひなの飼養試験では，異なる飼料が多数の特性（成熟日数，卵の大きさ，体重など）に与える影響の差の有意性が同時に検定される．研究のために選ばれた各特性は，実験者にとって特に興味あるものであろうが，いくつかの特性にあらわれる差は，他の特性にあらわれる差に付随的なものであるために，前者によって与えられる（後者と独立な）付加情報は無視できることがある．そのようなときには，前者の特性について観測値を得ることは不経済であろう．さらに，飼料によって直接的に影響される特性と，その変化が付随的なものにすぎない特性を見分けることができれば面白い．本節では，そのような目的に役立つ検定をいくつか述べよう．

測定された p 個の特性のうち，はじめの s 個の特性とは独立に，残りの $p - s$ 個の特性によって与えられる（付加）情報が有意であるか否かを調べよう．これは，残りの $p - s$

個の特性に有意な差があるか否かを調べることとは別である．この2つの問題のあいだの違いは，(8c. 4. 1) と (8c. 4. 2) の数学的定式化によって明らかになるであろう．前の問題に答えるために，観測値 $\mathbf{Y}_{s+1},\cdots,\mathbf{Y}_p$ の条件つき分布を考える．その場合のガウス＝マルコフ模型は，次のように書ける．

$$E(\mathbf{Y}_i)=\mathbf{X}\boldsymbol{\tau}_i+\gamma_{i1}\mathbf{Y}_1+\cdots+\gamma_{is}\mathbf{Y}_s$$
$$=\mathbf{X}\boldsymbol{\tau}_i+\mathscr{U}_1\boldsymbol{\gamma}_i,\qquad i=s+1,\cdots,p,\tag{8c. 4. 1}$$

ただし，$\mathbf{X}$ は前のモデル (8c. 1. 2) におけると同じ計画行列で，$\mathscr{U}_1=(\mathbf{Y}_1,\cdots,\mathbf{Y}_s)$，$\boldsymbol{\gamma}_i'=(\gamma_{i1},\cdots,\gamma_{is})$ である．もとの母数 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_p$ であらわすと，

$$\boldsymbol{\tau}_i=\boldsymbol{\beta}_i-\gamma_{i1}\boldsymbol{\beta}_1-\cdots-\gamma_{is}\boldsymbol{\beta}_s,\quad i=s+1,\cdots,p,\tag{8c. 4. 2}$$

であることがわかる．

条件づきでない期待値であらわすと，$\mathbf{Y}_{s+1},\cdots,\mathbf{Y}_p$ のガウス＝マルコフ模型は，

$$E(\mathbf{Y}_i)=\mathbf{X}\boldsymbol{\beta}_i,\qquad i=s+1,\cdots,p\tag{8c. 4. 3}$$

となる．他の特性とは無関係に，i 番目の特性に関して興味のある仮説を $\mathbf{H}'\boldsymbol{\beta}_i=0$ とする．条件づき期待値 (8c. 4. 1) に同じ仮説を適用すると，$\mathbf{H}'\boldsymbol{\tau}_i=0$ となる．これを $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_p$ であらわすと，

$$\mathbf{H}'\boldsymbol{\tau}_i=0=\mathbf{H}'\boldsymbol{\beta}_i-\gamma_{i1}\mathbf{H}'\boldsymbol{\beta}_1-\cdots-\gamma_{is}\mathbf{H}'\boldsymbol{\beta}_s$$

となるから，仮説 $\mathbf{H}'\boldsymbol{\beta}_i=0$ と $\mathbf{H}'\boldsymbol{\tau}_i=0$ とは異なる．偶然にも $\mathbf{H}'\boldsymbol{\beta}_i\neq 0$ であるが，$\mathbf{H}'\boldsymbol{\tau}_i=0$ となる，すなわち $\mathbf{H}'\boldsymbol{\beta}_i$ におけるずれが，量 $\mathbf{H}'\boldsymbol{\beta}_1,\cdots,\mathbf{H}'\boldsymbol{\beta}_s$ の（つまりこれらによって説明される）特殊な1次関数であるときがある．

仮説 $\mathbf{H}'\boldsymbol{\tau}_i=0$，$i=s+1,\cdots,p$ のことを，残りの $p-s$ 個の特性によって与えられる付加情報に関する帰無仮説と呼ぶことにしよう．(8c. 4. 1) の仮定の下で，そのような仮説を検定する問題を考えよう．この問題は，実は，4h. 2 で述べた，母数が2組あるときの，つまり付随的な変数にたいする共分散補正をするときの，最小2乗法の拡張にあたる．

模型 (8c. 4. 1) もまたガウス＝マルコフ型であるから，すなわち，より多くの母数が含まれる以外は (8c. 1. 2) と同じ形であることから，検定について新しい問題はなにも起らない．仮説からの偏差と残差による S. P. 行列を陽にあらわす式を導くには，$\mathbf{Y}_{s+1},\cdots,\mathbf{Y}_p$ の1次式を考え，4h. 2 の公式を用いて対応する平方和を求めるだけでよい．このとき，仮説 $\mathbf{H}'\boldsymbol{\tau}_i=0$ を検定するための S. P. 行列は，次のように求められる．まず，模型 (8c. 1. 2) を考え，$\mathbf{H}'\boldsymbol{\beta}_i=0$，$i=1,\cdots,p$ を検定するための S. P. 行列 $\mathbf{R}_0$ と $\mathbf{R}_1$ をふつうの方法で求める．これらの行列を分割して，

$$\mathbf{R}_0=\begin{pmatrix}\mathbf{R}_0(1,1) & \mathbf{R}_0(1,2)\\ \mathbf{R}_0(2,1) & \mathbf{R}_0(2,2)\end{pmatrix},\qquad \mathbf{R}_1=\begin{pmatrix}\mathbf{R}_1(1,1) & \mathbf{R}_1(1,2)\\ \mathbf{R}_1(2,1) & \mathbf{R}_1(2,2)\end{pmatrix}$$
$$\text{D.F.}=n-r\qquad\qquad \text{D.F.}=(n-r)+q$$

と書く．ただし引数1と2は，s 個と $p-s$ 個の変数を含む第1と第2の組を示し，$R(\mathbf{H})=q$ である．模型 (8c. 4. 1) の下での残差による S. P. 行列は，

$$\mathbf{R}_0(2\cdot1)=\mathbf{R}_0(2,2)-\mathbf{R}_0(2,1)[\mathbf{R}_0(1,1)]^{-1}\mathbf{R}_0(1,2)\tag{8c. 4. 4}$$
$$\text{D.F.}=n-r-s$$

で，"残差＋仮説からの偏差" による S. P. 行列は

$$\mathbf{R}_1(2\cdot1)=\mathbf{R}_1(2,2)-\mathbf{R}_1(2,1)[\mathbf{R}_1(1,1)]^{-1}\mathbf{R}_1(1,2)\tag{8c. 4. 5}$$
$$\text{D.F.}=(n-r-s)+q$$

となる．このとき，2つの行列は

$$\mathbf{R}_0(2\cdot1)\sim W_{p-s}(n-r-s,\Sigma_{2\cdot1}),$$
$$\mathbf{R}_1(2\cdot1)-\mathbf{R}_0(2\cdot1)\sim W_{p-s}(q,\Sigma_{2\cdot1})$$

となり，それらは互いに独立に分布し，帰無仮説が真のとき中心ウィシャート分布に従う．したがって，

$$|\mathbf{R}_0(2\cdot1)-\lambda\mathbf{R}_1(2\cdot1)|=0$$

の根に基づく任意の検定基準を適用できる．

検定基準が根の積

$$\Lambda_{(p-s)\cdot s}=\frac{|\mathbf{R}_0(2\cdot1)|}{|\mathbf{R}_1(2\cdot1)|}\tag{8c. 4. 6}$$

であるときには，その分布は $\Lambda(p-s,n-r-s,q)$ の分布となる．(8c. 4. 6) は行列 $\mathbf{R}_0,\mathbf{R}_1$ とその部分行列の行列式に陽に依存する，より簡単な形に書くことができる．すなわち

$$\frac{|\mathbf{R}_0(2\cdot1)|}{|\mathbf{R}_1(2\cdot1)|}=\frac{|\mathbf{R}_0|_p}{|\mathbf{R}_1|_p}\div\frac{|\mathbf{R}_0|_s}{|\mathbf{R}_1|_s}\tag{8c. 4. 7}$$

ただし，

$$|\mathbf{R}_0|_k=\begin{vmatrix}R_0(1,1) & \cdots & R_0(1,k)\\ \cdots & \cdots & \cdots\\ R_0(k,1) & \cdots & R_0(k,k)\end{vmatrix}$$

などである．(8c. 4. 7) は恒等式

$$|\mathbf{R}_0|_p=|\mathbf{R}_0|_s|\mathbf{R}_0(2\cdot1)|,\qquad |\mathbf{R}_1|_p=|\mathbf{R}_1|_s|\mathbf{R}_1(2\cdot1)|$$

から導かれる．

p 個の特性全部にたいして仮説 $\mathbf{H}'\boldsymbol{\beta}_i=0$ を同時に検定するための Λ 基準の分解として (8c. 4. 7) 式を次のように書くことができる．

$$\frac{|\mathbf{R}_0|_p}{|\mathbf{R}_1|_p}=\frac{|\mathbf{R}_0|_s}{|\mathbf{R}_1|_s}\times\frac{|\mathbf{R}_0(2\cdot1)|}{|\mathbf{R}_1(2\cdot1)|}$$
$$\Lambda_p=\Lambda_s\times\Lambda_{(p-s)\cdot s},\tag{8c. 4. 8}$$

ただし，Λ_s は最初の s 個の特性についての Λ 基準である．ここで導かれた $\Lambda_{(p-s)\cdot s}$ の

分布は，仮説 $\mathbf{H}'\tau_i = 0$, $i = s+1, \cdots, p$ が真であることのみに依存し，最初の s 個の特性に関係する母数 $\beta_1, \cdots, \beta_s$ の真の値が何であるかに無関係である（つまり，$i = 1, \cdots, s$ にたいして $\mathbf{H}'\beta_i = 0$ であるか否かには無関係である）ことに注意することが大切である．

$q=1$ のときの簡単化　仮説の D.F. $q=1$ のときには，比 $|\mathbf{R}_0|_p/|\mathbf{R}_1|_p$ はホテリングの T^2 を用いて次のように書けることが [(xii), **8b.2**] に示されている．

$$\frac{|\mathbf{R}_0|_p}{|\mathbf{R}_1|_p} = \frac{1}{1+T_p^2/k}, \qquad \text{ただし，} k \text{ は } \mathbf{R}_0 \text{ の D.F.}$$

帰無仮説 $\mathbf{H}'\beta_i = 0$ がすべての i について真ならば，

$$\frac{k-p+1}{p}\frac{T_p^2}{k} \sim F(p, k-p+1) \tag{8c.4.9}$$

となる．T_s^2 を，最初の s 個の特性に基づく（すなわち，単一仮説 $\mathbf{H}'\beta_i = 0$, $i = 1, \cdots, s$ を検定するための）ホテリング統計量であるとすると，

$$\frac{|\mathbf{R}_0|_s}{|\mathbf{R}_1|_s} = \frac{1}{1+T_s^2/k}$$

である．このとき，

$$\frac{|\mathbf{R}_0|_p}{|\mathbf{R}_1|_p} \div \frac{|\mathbf{R}_0|_s}{|\mathbf{R}_1|_s} = \frac{|\mathbf{R}_0(2\cdot1)|}{|\mathbf{R}_1(2\cdot1)|} = \frac{1}{1+T^2_{(p-s)\cdot s}/(k-s)}$$

と書くと，次の式

$$U_{(p-s)\cdot s} = \frac{T^2_{(p-s)\cdot s}}{k-s} = \frac{T_p^2 - T_s^2}{k+T_s^2} \tag{8c.4.10}$$

を得，これから，$i = 1, \cdots, s$ にたいして，$\mathbf{H}'\beta_i$ が何であっても，付加情報がまったくないという帰無仮説 $\mathbf{H}'\tau_i = 0$, $i = s+1, \cdots, p$ が真であるときの，$(p-s, k-p+1)$D.F. をもつ分散比

$$\frac{k-p+1}{p-s}\cdot\frac{T_p^2 - T_s^2}{k+T_s^2} \tag{8c.4.11}$$

が導ける．

8c.5　Λ の 分 布

Λ 検定の適用には，次の表記法が一般に用いられる．

表 8c.5α　分散共分散分析（p 特性）

変動因	D.F.	S.P. 行列
仮説からの偏差	q	$\mathbf{R}_1 - \mathbf{R}_0$
残　差(誤差)	$t-q$	$\mathbf{R}_0$
全体	t	$\mathbf{R}_1$
$\Lambda = \lvert\mathbf{R}_0\rvert \div \lvert\mathbf{R}_1\rvert$		

パラメータ $(p, t-q, q)$ をもつ Λ は，パラメータ

$$\left(\frac{t-q-p+1}{2}, \frac{q}{2}\right), \quad \cdots, \quad \left(\frac{t-q}{2}, \frac{q}{2}\right) \tag{8c.5.1}$$

をもつ独立なベータ変量の積として分布することは，[(xi), **8b.2**] で述べた．いくつかの特別な場合，この積の分布は，変換により，正確な分散比〔F〕に帰着できる (Nair(1939), Wilks(1932))．その結果を表 8c.5β にまとめる．

表 8c.5β　特別な場合の正確な分散比〔F〕

p, q の値	分散比	D.F.
$q=1$, p 任意	$\frac{1-\Lambda}{\Lambda}\frac{t-p}{p}$	p, $t-p$
$q=2$, p 任意	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\frac{t-p-1}{p}$	$2p$, $2(t-p-1)$
$p=1$, q 任意	$\frac{1-\Lambda}{\Lambda}\frac{t-q}{q}$	q, $t-q$
$p=2$, q 任意	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\frac{t-q-1}{q}$	$2q$, $2(t-q-1)$

p, q の他の値にたいしては，$\Lambda(p, t-q, q)$ のよい近似がある．Bartlett(1947) によるものは

$$-m\log_e\Lambda = -\left(t - \frac{p+q+1}{2}\right)\log_e\Lambda \tag{8c.5.2}$$

が，自由度 pq をもつ中心 χ^2 分布をするものとして用いる．著者 (Rao (1951d)) によるものは，一層よい近似を与え，

$$R = \frac{1-\Lambda^{1/s}}{\Lambda^{1/s}}\frac{ms-2\lambda}{pq} \tag{8c.5.3}$$

を，自由度 pq と $ms-2\lambda$ の分散比〔F〕として用いる．ただし

$$m = t - \frac{p+q+1}{2}, \quad s = \sqrt{\frac{p^2q^2-4}{p^2+q^2-5}}, \quad \lambda = \frac{pq-2}{4} \tag{8c.5.4}$$

である．$ms-2\lambda$ は整数である必要はないことを注意しておこう．それでも，分散比の有意点を見るのに困難を生じない．それに適合する値は，自由度 $[ms-2\lambda]$ と $[ms-2\lambda]+1$ にたいする有意点のあいだにある．実用上は，分母の自由度として，$ms-2\lambda$ よりは小さいかまたは等しいもののなかで最大の整数 $[ms-2\lambda]$ を用いるのが安全である．この分布関数の他の近似および漸近展開については，Rao(1948b, 1951d) および J. Roy (1951) を参照せよ．Λ の正確な分布に関する研究については，Rao(1972a) を見よ．

8c.6　次元数の検定（構造関係）

次元数に関する仮説　k 個の p 変量正規母集団 $N_p(\mu_1, \Sigma), \cdots, N_p(\mu_k, \Sigma)$ を考える．

この平均値の成分の間になんらかの構造関係があるかどうかを調べることに興味が生じるかも知れない (Fisher (1939))．線形の構造関係は

$$\mathbf{H}\boldsymbol{\mu}_i = \boldsymbol{\xi}, \qquad i = 1, \cdots, k \tag{8c. 6. 1}$$

という形であらわせる．ただし，$\mathbf{H}$ は $s \times p$ 行列で，$\boldsymbol{\xi}$ は $s \times 1$ ベクトルである．$\mathbf{H}$ と $\boldsymbol{\xi}$ は未知であるが，式 (8c. 6. 1) ですべての i について一定である．仮説 (8c. 6. 1) から，E_p における点 $\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_k$ は，実際には，$r = p - s$ 次元空間の上にあることがわかる．事実，$\boldsymbol{\mu}$ が E_p の1点をあらわすとき，式 $\mathbf{H}\boldsymbol{\mu} = \boldsymbol{\xi}$ は $p - s$ 次元空間をあらわす式になる．仮説 (8c. 6. 1) は，$r < k - 1$ のときのみ興味があることに注意しておこう．

データ　$N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ から，それぞれ N_i 個の観測値がとられたとし，

$$\overline{\mathbf{U}}_i \quad と \quad (S_{rs}^{(i)})$$

を，この N_i 個 $(i = 1, \cdots, k)$ の観測値に基づく標本平均と，補正済 S. P. 行列とする．

群（母集団）間および群内の分散共分散分析は，次のようになる．

変動因	D. F.	S. P. 行列
群　間	$k-1$	$\mathbf{B} = \sum N_i \overline{\mathbf{U}}_i \overline{\mathbf{U}}_i' - (\sum N_i)\overline{\mathbf{U}}\overline{\mathbf{U}}'$
群　内	$\sum N_i - k$	$\mathbf{W} = (\sum_i S_{rs}^{(i)})$
全　体	$\sum N_i - 1$	$\mathbf{T} = (S_{rs})$

ただし，$\overline{\mathbf{U}}$ は総平均，(S_{rs}) は $N_1 + \cdots + N_k$ 個の観測値からなる全標本から求めた補正済 S. P. 行列である．

検定基準　$\boldsymbol{\Sigma}$ は既知で正則とする．このとき，観測平均値の密度は，定数を除くと，

$$L(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_k) = \exp\left[-\frac{1}{2}\sum_{i=1}^{k} N_i(\overline{\mathbf{U}}_i - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{U}} - \boldsymbol{\mu}_i)\right] \tag{8c. 6. 2}$$

となる．(8c. 6. 2) を $\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_k$ の関数と考え，仮説 H_0：$\boldsymbol{\mu}_i$ は r 次元空間上にある，を検定するための尤度比基準，

$$\chi^2 = -2\log_e \frac{\sup_{H_0} L(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_k)}{\sup L(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_k)} \tag{8c. 6. 3}$$

を導く．

(8c. 6. 3) の分子では，$\boldsymbol{\mu}_i$，$i = 1, \cdots, k$ は r 次元空間内にあるという制約の下で，上限が求められるが，分母では $\boldsymbol{\mu}_i$ に制約はない．尤度比基準の一般論から，統計量 (8c. 6. 3) の漸近分布は，標本の大きさ $N_1, \cdots, N_k$ が大きいとき，pk 個の母数に仮説 H_0 によって付与される制約数に等しい D. F. をもつ χ^2 分布になる． この数は，r 次元空間が $r+1$ 個の点で規定され，その空間上の任意の点は係数の和が1になるような $r+1$ 個の点の1次結合によってあらわされることに注意すれば，すぐに計算できる．各点は，p 個の座標，つまり p 個の母数によってあらわされる．それゆえ，自由な（任意に定められる）母数の

数は，$p(r+1) + (k-r-1)r$ となり，制約数は，

$$pk - p(r+1) - (k-r-1)r = (p-r)(k-r-1) \tag{8c. 6. 4}$$

となる．したがって，χ^2 の D. F. は $(p-r)(k-r-1)$ である．

さて，$\boldsymbol{\mu}_i$ に制約がないときは，$\sup \log L(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_k) = 0$ である，そして本節で後に示すように

$$\sup_{H_0} \log L = -\frac{1}{2}(\lambda_{r+1} + \cdots + \lambda_p) \tag{8c. 6. 5}$$

である．ただし，$\lambda_{r+1}, \cdots, \lambda_p$ は固有方程式 $|\mathbf{B} - \lambda\boldsymbol{\Sigma}| = 0$ の最小の $p - r$ 個の固有値である．したがって，(8c. 6. 3) の χ^2 は，

$$\chi^2 = \lambda_{r+1} + \cdots + \lambda_p, \qquad \text{D. F.} = (p-r)(k-r-1) \tag{8c. 6. 6}$$

となる．$n = \sum N_i - k$ が大きく，$\boldsymbol{\Sigma}$ が未知なら，推定値 $n^{-1}\mathbf{W}$ を $\boldsymbol{\Sigma}$ のかわりに用いることができる．その場合，(8c. 6. 6) の χ^2 検定は，近似的ではあるが，やはり有効であろう．$\mathbf{W}$ を n で割らないで，固有方程式

$$|\mathbf{B} - \nu\mathbf{W}| = 0 \tag{8c. 6. 7}$$

を考えるなら，

$$\chi^2 = n(\nu_{r+1} + \cdots + \nu_p), \qquad \text{D. F.} = (p-r)(k-r-1)$$

となる，ただし，$\nu_{r+1}, \cdots, \nu_p$ は方程式 (8c. 6. 7) の最小の $p - r$ 個の根である．

多くの場合，平均値（ベクトル）のつくる図形の次元を定めることが問題となる．これには，根とそれに対応する D. F. を大きい方から順に並べ，それらを上の方に累積すると，次のようにして接近できる．

根	D. F.	累積 根	累積 D. F.
λ_1	$p+k-2$	$\chi_0^2 = \lambda_1 + \cdots + \lambda_p$	$p(k-1)$
λ_2	$p+k-4$	$\chi_1^2 = \lambda_2 + \cdots + \lambda_p$	$(p-1)(k-2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
λ_{r+1}	$p+k-2r-2$	$\chi_r^2 = \lambda_{r+1} + \cdots + \lambda_p$	$(p-r)(k-r-1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

もし，$\chi_r^2, \chi_{r+1}^2, \cdots$ はその D. F. に比べて有意に小さいが，$\chi_{r-1}^2, \chi_{r-2}^2, \cdots$ は大きいなら，その次元数は r であると推測できる．

上に述べた手順からの類推で，

$$|\mathbf{W} - \theta(\mathbf{B} + \mathbf{W})| = 0 \tag{8c. 6. 8}$$

の小さい方の根の積に基づく別の検定基準を考えることができる．ただし，λ と θ には，式

$$\left(1 + \frac{\lambda}{n}\right) = \frac{1}{\theta}$$

で結ばれる関係がある．$\lambda_{r+1}+\cdots+\lambda_p$ に対応する統計量は，

$$-n\log(\theta_{r+1}\cdots\theta_p)=-n(\log\theta_{r+1}+\cdots+\log\theta_p)$$

である．Bartlett (1947b) は，

$$-\left(\sum N_i-1-\frac{p+k}{2}\right)(\log\theta_{r+1}+\cdots+\log\theta_p) \qquad (8c.6.9)$$

が，近似的に $(p-r)(k-r-1)$D.F. の χ^2 分布をすることを示唆した．

(8c.6.5) の証明　問題は

$$\min_{H_0}\sum N_i(\bar{\mathbf{U}}_i-\mu_i)'\Sigma^{-1}(\bar{\mathbf{U}}_i-\mu_i)=\lambda_{r+1}+\cdots+\lambda_p \qquad (8c.6.10)$$

を示すことである．ただし，$\lambda_1\geqslant\cdots\geqslant\lambda_p$ は

$$|\mathbf{B}-\lambda\Sigma|=0 \qquad (8c.6.11)$$

の根とし，行列 $\mathbf{B}$ は $\sum N_i\bar{\mathbf{U}}_i\bar{\mathbf{U}}_i'-(\sum N_i)\bar{\mathbf{U}}\bar{\mathbf{U}}'$ である．$\mathbf{Z}_i=\Sigma^{-1/2}\bar{\mathbf{U}}_i$，$\zeta_i=\Sigma^{-1/2}\mu_i$ とおくと，式 (8c.6.10) は

$$\min_{H_0}\sum_{i=1}^k N_i(\mathbf{Z}_i-\zeta_i)'(\mathbf{Z}_i-\zeta_i) \qquad (8c.6.12)$$

と同じになる．ただし，H_0 は ζ_i に関する同じ仮説，すなわち，ζ_i は r 次元空間内にあることである．1つの r 次元空間は，1点 $\mathbf{X}_0$ と r 個の正規直交系 $\mathbf{X}_1,\cdots,\mathbf{X}_r$ によって決まる．そしてこの空間上の1点 ζ_i は，

$$\zeta_i=\mathbf{X}_0+C_{i1}\mathbf{X}_1+\cdots+C_{ir}\mathbf{X}_r \qquad (8c.6.13)$$

という形にあらわせる．(8c.6.12) の i 番目の項は，乗数 N_i を除くと，

$$(\mathbf{Z}_i-\mathbf{X}_0-\sum C_{ij}\mathbf{X}_j)'(\mathbf{Z}_i-\mathbf{X}_0-\sum C_{ij}\mathbf{X}_j) \qquad (8c.6.14)$$

と書ける．所与の $\mathbf{X}_0,\mathbf{X}_1,\cdots,\mathbf{X}_r$ にたいして，$\mathbf{X}_1,\cdots,\mathbf{X}_r$ が正規直交系であることを用いると，式 (8c.6.14) の最小値は（通常の回帰理論と残差平方和の式を用いて），

$$(\mathbf{Z}_i-\mathbf{X}_0)'(\mathbf{Z}_i-\mathbf{X}_0)-[(\mathbf{Z}_i-\mathbf{X}_0)'\mathbf{X}_1]^2-\cdots-[(\mathbf{Z}_i-\mathbf{X}_0)'\mathbf{X}_r]^2 \qquad (8c.6.15)$$

となる．これに N_i をかけ，i について和をとると，$\mathbf{X}_0,\mathbf{X}_1,\cdots,\mathbf{X}_r$ に関して最小にすべき式は，

$$\sum_{i=1}^k N_i(\mathbf{Z}_i-\mathbf{X}_0)'(\mathbf{Z}_i-\mathbf{X}_0)-\sum_{i=1}^k\sum_{j=1}^r N_i[(\mathbf{Z}_i-\mathbf{X}_0)'\mathbf{X}_j]^2 \qquad (8c.6.16)$$

となることがわかる．$\bar{\mathbf{Z}}=\sum N_i\mathbf{Z}_i/\sum N_i$ を導入することによって，(8c.6.16) は，

$$\sum_{i=1}^k N_i(\mathbf{Z}_i-\bar{\mathbf{Z}})'(\mathbf{Z}_i-\bar{\mathbf{Z}})-\sum_{i=1}^k\sum_{j=1}^r N_i[(\mathbf{Z}_i-\bar{\mathbf{Z}})'\mathbf{X}_j]^2$$

$$+N(\bar{\mathbf{Z}}-\mathbf{X}_0)'(\bar{\mathbf{Z}}-\mathbf{X}_0)-N\sum_{j=1}^r[(\bar{\mathbf{Z}}-\mathbf{X}_0)'\mathbf{X}_j]^2 \qquad (8c.6.17)$$

と書ける．$\bar{\mathbf{Z}}-\mathbf{X}_0$ を含む式は非負であるから，所与の $\mathbf{X}_1,\cdots,\mathbf{X}_r$ にたいして，$\mathbf{X}_0=\bar{\mathbf{Z}}$ のとき，(8c.6.17) は最小となる．このとき，$\mathbf{X}_1,\cdots,\mathbf{X}_r$ に関して最小にすべき式は，

$$\sum_{i=1}^k N_i(\mathbf{Z}_i-\bar{\mathbf{Z}})'(\mathbf{Z}_i-\bar{\mathbf{Z}})-\sum_{i=1}^k\sum_{j=1}^r N_i[(\mathbf{Z}_i-\bar{\mathbf{Z}})'\mathbf{X}_j]^2 \qquad (8c.6.18)$$

となる．この第1項は $\mathbf{X}_i$ とは独立であるから，第2項を $\mathbf{X}_1,\cdots,\mathbf{X}_r$ に関して最大にする必要がある．ここで，

$$\sum_{i=1}^k\sum_{j=1}^r N_i[\mathbf{X}_j'(\mathbf{Z}_i-\bar{\mathbf{Z}})]^2=\sum_{j=1}^r\mathbf{X}_j'\left[\sum_{i=1}^k N_i(\mathbf{Z}_i-\bar{\mathbf{Z}})(\mathbf{Z}_i-\bar{\mathbf{Z}})'\right]\mathbf{X}_j=\sum_{j=1}^r\mathbf{X}_j'\mathbf{C}\mathbf{X}_j$$

で，$\mathbf{C}=\sum N_i(\mathbf{Z}_i-\bar{\mathbf{Z}})(\mathbf{Z}_i-\bar{\mathbf{Z}})'$ である．固有方程式 $|\mathbf{C}-\lambda\mathbf{I}|=0$ の固有ベクトルを $\mathbf{P}_1,\mathbf{P}_2,\cdots,\mathbf{P}_p$，固有値を〔大きさの順に〕$\lambda_1,\cdots,\lambda_p$ とすると，(1f.2.8) で示したように，

$$\max_{\mathbf{X}_1,\cdots,\mathbf{X}_r}\sum_1^r\mathbf{X}_j'\mathbf{C}\mathbf{X}_j=\mathbf{P}_1'\mathbf{C}\mathbf{P}_1+\cdots+\mathbf{P}_r'\mathbf{C}\mathbf{P}_r$$
$$=\lambda_1+\cdots+\lambda_r$$

を得る．しかしながら，$\lambda_1+\cdots+\lambda_p=\mathrm{trace}\,\mathbf{C}=\sum N_i(\mathbf{Z}_i-\bar{\mathbf{Z}})'(\mathbf{Z}_i-\bar{\mathbf{Z}})$ であるから (8c.6.18) の最小値は $\lambda_{r+1}+\cdots+\lambda_p$ となる．

$$\mathbf{Z}_i=\Sigma^{-1/2}\bar{\mathbf{U}}_i,\qquad \bar{\mathbf{Z}}=\Sigma^{-1/2}\bar{\mathbf{U}}$$

であるから，

$$\sum N_i(\mathbf{Z}_i-\bar{\mathbf{Z}})(\mathbf{Z}_i-\bar{\mathbf{Z}})'=\Sigma^{-1/2}[\sum N_i(\bar{\mathbf{U}}_i-\bar{\mathbf{U}})(\bar{\mathbf{U}}_i-\bar{\mathbf{U}})']\Sigma^{-1/2}$$
$$\mathbf{C}=\Sigma^{-1/2}\mathbf{B}\Sigma^{-1/2}$$

である．したがって，

$$|\mathbf{C}-\lambda\mathbf{I}|=0=|\mathbf{B}-\lambda\Sigma|$$

となる．よって，$\lambda_1,\cdots,\lambda_p$ は固有方程式 $|\mathbf{B}-\lambda\Sigma|=0$ の根となり，必要な結果が証明されたことになる．

8c.7　構造母数をもつときの分散共分散分析（成長モデル）

8c.1 で導入した一般多変量線形模型

$$(\mathscr{U},\mathbf{X}\mathbf{B},\Sigma\otimes\mathbf{I}_n) \qquad (8c.7.1)$$

を考えよう．ここで，$\mathscr{U}$ は次数 $n\times p$，$\mathbf{X}$ は次数 $n\times m$，$\mathbf{B}$ は次数 $m\times p$，Σ は次数 $p\times p$ の行列である．8c.2—8c.6 の解析では，$\mathbf{B}$ の mp 個の母数は自由と考えた．ここでは，$\mathbf{B}$ の各行の母数の間に線形制約式を導入しよう．それは，

$$\mathbf{B}=\Theta_0+\Theta\mathbf{H}$$

という形の関係を考えることによって導かれる．ただし，Θ は自由な母数より成る $m\times k$ 行列で，Θ_0 は次数 $m\times p$，$\mathbf{H}$ は次数 $k\times p$ でともに所与の行列である．そのような場合，モデル (8c.7.1) は

$$(\mathscr{U}-\mathbf{X}\Theta_0,\mathbf{X}\Theta\mathbf{H},\Sigma\otimes\mathbf{I}_n) \qquad (8c.7.2)$$

と書け，成長データに関する多数の研究 (Rao (1959a, 1961g, 1965b)) で用いられている．

モデル (8c. 7. 2) の一般式は Potthoff and Roy(1964) による．

モデルの簡約化　$R(\mathbf{H})=k$ とし，$\mathbf{Z}$ は階数 $p-k$ の $p\times(p-k)$ 行列で $\mathbf{HZ}=0$ とする．さらに，Σ_0 を Σ のアプリオリな値とし，正則であると仮定する．[われわれの理論構成上は，Σ_0 は任意の p. d. 行列であればよい．アプリオリな値が得られないときは，それを $\mathbf{I}$ とすればよい．]

$$\mathbf{W}_1=(\mathscr{U}-\mathbf{X}\Theta_0)\Sigma_0^{-1}\mathbf{H}'(\mathbf{H}\Sigma_0^{-1}\mathbf{H}')^{-1}$$
$$\mathbf{W}_2=(\mathscr{U}-\mathbf{X}\Theta_0)\mathbf{Z} \qquad (8c. 7. 3)$$

を考える．このとき，$E(\mathbf{W}_1)=\mathbf{X}\Theta$，$E(\mathbf{W}_2)=0$ となり，モデル (8c. 7. 1) を変数 $\mathbf{W}_1$，$\mathbf{W}_2$ であらわすと

$$[(\mathbf{W}_1 \vdots \mathbf{W}_2),(\mathbf{X}\Theta \vdots 0),\Lambda\otimes\mathbf{I}_n] \qquad (8c. 7. 4)$$

となる．ここで，Λ は，$(\mathbf{W}_1 \vdots \mathbf{W}_2)$ における変数の分散共分散行列である．$E(\mathbf{W}_2)=0$ であるから，$\mathbf{W}_2$ を付随変数についての観測値として用い，モデル (8c. 7. 4) を

$$(\mathbf{W}_1,\mathbf{X}\Theta+\mathbf{W}_2\Gamma,\Lambda_1\otimes\mathbf{I}_n) \qquad (8c. 7. 5)$$

と書くことができる，ここで，Γ は (8c. 4. 1) におけると同様に付加的な回帰母数である．このようにして，モデル (8c. 7. 1) は自由な母数 Θ,Γ をもつモデルに簡約化される．このとき，共分散補正を用いる解析が適用され，新しい問題はなにも生じない．

$R(\mathbf{H})=r<k$ のときは，次のようにする．正則行列 $(\mathbf{H}_1 \vdots \mathbf{H}_2)$ をつくり，$\mathbf{HH}_2=0$ で，かつ，$\mathbf{H}_1$ の列は，$\mathbf{H}$ の行によって生成されるベクトル空間の基底を形成するようにする．このとき，変換

$$\mathbf{W}_1=(\mathscr{U}-\mathbf{X}\Theta_0)\Sigma_0^{-1}\mathbf{H}_1$$
$$\mathbf{W}_2=(\mathscr{U}-\mathbf{X}\Theta_0)\mathbf{H}_2 \qquad (8c. 7. 6)$$

を施し，$E(\mathbf{W}_1)=\mathbf{X}\Theta\mathbf{H}\Sigma_0^{-1}\mathbf{H}_1$，$E(\mathbf{W}_2)=0$ となるようにせよ．こんどは $R(\mathbf{H}\Sigma_0^{-1}\mathbf{H}_1)=r$ となるから，$\Theta\mathbf{H}\Sigma_0^{-1}\mathbf{H}_1$ は，自由な母数から成る $m\times r$ 行列 Φ におきかえることができる．このとき，モデル (8c. 7. 1) は

$$(\mathbf{W}_1,\mathbf{X}\Phi+\mathbf{W}_2\Gamma,\Lambda_1\otimes\mathbf{I}_n) \qquad (8c. 7. 7)$$

に簡約化され，ここでも，(8c. 7. 5) と同じ形のものとなる．

実際には，付随的観測値 $\mathbf{W}_2$ の有用性を吟味するという問題がある．Σ_0 が Σ の良い近似であるときは，$\mathbf{W}_2$ を省き，Θ または Φ についての推測をするのに，モデル $(\mathbf{W}_1,\mathbf{X}\Theta,\Lambda_1\otimes\mathbf{I}_n)$，または $(\mathbf{W}_1,\mathbf{X}\Phi,\Lambda_1\otimes\mathbf{I}_n)$ を考えればよい．しかしながら，共分散補正の有用性は，経験的な立証を要する問題である (Rao (1965b))．

8d　多変量検定のいくつかの適用

8d. 1　指定された平均値の検定

28本の樹について，北 (N)，東 (E)，南 (S)，および西 (W) 方向に穴をあけてコルク層の厚さを測定し，観測値を得た．問題はコルク層の厚さがすべての方向で同一であるかどうかを確かめることである．3つの特性（対比）が考えられる．すなわち

$$U_1=(N+S)-(E+W),\quad U_2=N-S,\quad U_3=E-W, \qquad (8d. 1. 1)$$

であり，U_1,U_2,U_3 の期待値が0であるかどうかを検定することになる．対比 (8d. 1. 1) のかわりに，いろいろな比較が選べるであろう．しかし，(8d. 1. 1) を選んだのは，コルク層の厚さは北向きと南向きでは同じになり，東向きと西向きでも同じであろうと考えたからである．したがって，最大の偏りは $(N+S)-(E+W)$ にあることが期待される．対比 (8d. 1. 1) は上述の仮説を直接検定できるように選んだ．

標本平均と標本分散共分散行列（すなわち，補正済平方和と積和を 27 D. F. で割った値）は

$$\bar{U}_1=8.8571,\quad \bar{U}_2=0.8571,\quad \bar{U}_3=1.0000$$

$$(s_{ij})=\begin{pmatrix}128.7200 & -21.0211 & -26.9259\\ & 63.5344 & 21.3544\\ & & 96.5761\end{pmatrix}$$

である．行列 (s_{ij}) の逆行列を (s^{ij}) であらわす．

ホテリングの T^2 が指定された平均値の検定に適用でき，その一般式は [検定における特性数 p，標本の大きさ N，推定平均値 $\bar{U}_1,\cdots,\bar{U}_p$ および $(N-1)$ D. F. の推定分散共分散行列 (s_{ij})，期待値 $\xi_1,\xi_2,\cdots$ にたいしては]，

$$T_p^2=N\sum\sum s^{ij}(\bar{U}_i-\xi_i)(\bar{U}_j-\xi_j) \qquad (8d. 1. 2)$$

である．N が大きければ，T_p^2 は p D. F. の χ^2 分布となる．小さな N にたいする適当な手順は，D. F. が p と $(N-p)$ の分散比 F 統計量として

$$F=\frac{N-p}{(N-1)p}T_p^2 \qquad (8d. 1. 3)$$

を用いることである．この例では $p=3$，$N=28$，および $\xi_i=0$，$i=1,2,3$ である．

$$\sum\sum s^{ij}(\bar{U}_i-\xi_i)(\bar{U}_j-\xi_j) \qquad (8d. 1. 4)$$

2次形式 (8d. 1. 4) を計算する単純な方法は以下の通りである．次の行列

分散共分散行列 (s_{ij})			偏差 $(\bar{U}_i-\xi_i)$
128.7200	−21.0211	−26.9259	8.8571
	63.5344	21.3544	0.8571
		96.5761	1.0000
			0

から始めて，枢軸変換によって掃き出す．その最後の値の符号を変えると，2次形式(8d.1.4) の値となる．この例における値は 0.7541 となることがわかる．これから

$$T_p^2 = N(0.7541) = 21.1148$$

$$F = \frac{N-p}{(N-1)p}T_p^2 = \frac{(28-3)\times 28}{27\times 3}(0.7541) = 6.5169$$

が得られ，この値 6.5169 は，3 と 25 D. F. の分散比であり，1 %水準で有意となる．

それぞれの $\overline{U}_i$ について，その残りを無視して，個別に検定をして見よう．それぞれは，1 と 27 D. F. の分散比になるような T_1^2 である．

$$N+S-E-W \text{ にたいする } F = T_1^2 = \frac{N(\overline{U}_1-\xi_1)^2}{s_{11}} = \frac{28(8.8571)^2}{128.7200} = 17.0646$$

$$N-S \text{ にたいする } F = T_1^2 = \frac{N(\overline{U}_2-\xi_2)^2}{s_{22}} = \frac{28(0.8571)^2}{63.5344} < 1$$

$$E-W \text{ にたいする } F = T_1^2 = \frac{N(\overline{U}_3-\xi_3)^2}{s_{33}} = \frac{28(1.0000)^2}{96.5761} < 1$$

すなわち，$N+S-E-W$ にたいする F は大きく，残りは小さい．これは，ここで考えた仮説が真であることを示している．$N-S$ および $E-W$ にたいする F 値が大きい場合を考えて見よう．そのときは，$N-S$ および $E-W$ における大きな差は，単に $N+S-E-W$ との相関によるのかどうかという疑問が生じる．この問題は **8c.4** で展開した付加情報にたいする検定によって調べることができる．仮説の D. F. は 1 であるので，式 (8c.4.11) が使える．s 個の特性とは独立に，$p-s$ 個の特性によって与えられる付加情報の有意性の検定にたいする基準は，式 (8c.4.11) で k のかわりに $N-1$ とおいた分散比

$$F = \frac{N-p}{p-s}\frac{T_p^2-T_s^2}{(N-1)+T_s^2} \tag{8d.1.5}$$

である ((8c.4.11) 参照)．この例では，$p=3$，$s=1$ で，$N+S-E-W$ にたいする T_1^2 の値は 17.0646 である．(8d.1.5) に代入すれば，

$$F = \frac{25}{2}\frac{21.1148-17.0646}{44.0646} = 1.1489,$$

となり，これは 2 と 25 D. F. の分散比としては小さい値である．

注意　T_p^2 はいろいろな方法で計算できるので，いくつかの別の公式を示す．$(S_{ij}) = \mathbf{S}$ は補正済平方和積和行列，$(s_{ij}) = \mathbf{s} = (N-1)^{-1}\mathbf{S}$，および $\mathbf{d}' = (\overline{U}_1-\xi_1, \cdots, \overline{U}_p-\xi_p)$ とすると，

$$\begin{aligned}T_p^2 &= N\sum\sum s^{ij}(\overline{U}_i-\xi_i)(\overline{U}_j-\xi_j)\\ &= N(N-1)\sum\sum S^{ij}(\overline{U}_i-\xi_i)(\overline{U}_j-\xi_j)\\ &= -N\left|\begin{array}{c:c}\mathbf{s} & \mathbf{d}\\ \hdashline \mathbf{d}' & 0\end{array}\right| \div |\mathbf{s}| = N\left\{\frac{|\mathbf{s}+\mathbf{dd}'|}{|\mathbf{s}|}-1\right\} = \frac{|\mathbf{s}+N\mathbf{dd}'|}{|\mathbf{s}|}-1\end{aligned}$$

$$\begin{aligned}&= -N(N-1)\left|\begin{array}{c:c}\mathbf{S} & \mathbf{d}\\ \hdashline \mathbf{d}' & 0\end{array}\right| \div |\mathbf{S}| = N(N-1)\left\{\frac{|\mathbf{S}+\mathbf{dd}'|}{|\mathbf{S}|}-1\right\}\\ &= (N-1)\left\{\frac{|\mathbf{S}+N\mathbf{dd}'|}{|\mathbf{S}|}-1\right\}.\end{aligned}$$

となる．

8d.2　平均値の所与の構造にたいする検定

N 個の独立な観測値が p 変量母集団から得られる場合を考える．**8d.1** におけると同様に，$\overline{U}_1, \cdots, \overline{U}_p$ を標本平均，(s_{ij}) を $(N-1)$ D. F. の標本（推定）分散共分散行列とする．そのほかに，理論平均値を $\mu_1, \cdots, \mu_p$ とする．興味ある仮説のいくつかを次に示す．

1.　すべての i にたいして，$\mu_i = \mu$（すなわち，すべての特性は同一の平均値をもつ）．
2.　$p = 2r$ のとき $\mu_i = \mu_{r+i}$，$i = 1, \cdots, r$．このような仮説は，生物体の左右両側から，いくつかの特性値を計測して，対称性に関して検定したいときにでてくる．
3.　$\mu_i = \alpha + \beta_1 t_i + \cdots + \beta_r t_i^r$，ただし，$\alpha, \beta_1, \cdots, \beta_r$ は未知の母数で，t_i は既知とする．たとえば，t_i を i 番目の測定値が採られる時間とし，平均値にたいして t_i の多項式による傾向曲線をあてはめる場合がこの仮説に対応する．

このような仮説は，すべて統計量

$$R_0^2 = \min_{H_0}\sum\sum s^{ij}(\overline{U}_i-\mu_i)(\overline{U}_j-\mu_j) \tag{8d.2.1}$$

によって検定できる．ただし，H_0 は μ_i，$i = 1, \cdots, p$ に関する制約をあらわす．(8d.2.1) に基づく分散比は，q と $(N-q)$ D. F. の

$$F = \frac{N-q}{q}\cdot\frac{N}{N-1}R_0^2 \tag{8d.2.2}$$

である．ただし，q は $\mu_1, \cdots, \mu_p$ に関する独立な制約の数である．したがって，(1) の場合には，$q = p-1$ となり，(2) では $q = r$，そして (3) では $q = p-(r+1)$，等々である．

平均値に関する複合仮説を含む (1) の場合は，**8d.1** で説明した例と同じである．この問題は確率変数の p 個の成分の $p-1$ 個の対比を考えることによって，平均値に関する単純仮説の検定に帰着できることを調べた．このことは，最初の p 個の確率変数を $p-1$ 個の確率変数に変換し，新しい確率変数として取り扱うことを意味した．ここでは，もとの確率変数の推定平均と推定分散共分散を用いて，ホテリングの T^2 を陽に表現してみよう．$\overline{U}_1, \cdots, \overline{U}_p$ は平均値，(s_{ij}) は $(N-1)$D. F. の標本分散 共分散行列とする．仮説 $\mu_1 = \cdots = \mu_p$ を検定するための (8d.2.1) の R_0^2 統計量は

$$R_0^2 = \min_{\mu}\sum\sum s^{ij}(\overline{U}_i-\mu)(\overline{U}_j-\mu)$$

であり，これは，読者が簡単に証明できるように，陽には次のように書ける．

$$\sum\sum s^{ij}\bar{U}_i\bar{U}_j - \frac{[\sum\sum s^{ij}(\bar{U}_i+\bar{U}_j)]^2}{4\sum\sum s^{ij}}.$$

他の場合の陽な表現については，Rao (1959a) の例を参照せよ．

8d.3　2つの母集団の平均値の差の検定

N_1, N_2 個の標本が2つの p 変量母集団から得られるとする．この2群の標本平均を

$$\bar{U}_{11}, \cdots, \bar{U}_{p1} \quad \text{および} \quad \bar{U}_{12}, \cdots, \bar{U}_{p2}$$

とあらわし，群内の補正済 S.P. 行列を

$$(S_{ij}^{(1)}), \qquad (S_{ij}^{(2)})$$
$$\text{D.F.} = N_1-1, \qquad \text{D.F.} = N_2-1$$

とし，併合した行列を

$$(S_{ij}^{(0)}) = (S_{ij}^{(1)} + S_{ij}^{(2)}), \qquad \text{D.F.} = N_1+N_2-2$$

とあらわす．

母集団平均の差が（すべての特性に関して同時に）ゼロであるという仮説を検定するために，マハラノビスの D^2 検定と同等なホテリングの検定を用いることができる．マハラノビスの D^2 は

$$D_p^2 = \sum\sum s^{ij}d_id_j \qquad (8d.3.1)$$

で定義される．ただし，$d_i = \bar{U}_{i2} - \bar{U}_{i1}$，$(s^{ij}) = (s_{ij})^{-1}$ で $s_{ij} = S_{ij}^{(0)}/(N_1+N_2-2)$ である．D_p^2 に基づく分散比は，p と (N_1+N_2-p-1)D.F. で

$$F = \frac{N_1+N_2-p-1}{p}\cdot\frac{N_1N_2}{(N_1+N_2)(N_1+N_2-2)}\cdot D_p^2 \qquad (8d.3.2)$$

である．表 8d.3α に示す例について考察しよう．

表 8d.3α　2つの種からの50ずつの観測値に基づく標本平均（出典：Fisher (1936)）

特　性	*Iris Versicolor*	*Iris Setosa*	差
がく片長	5.936	5.006	0.930
がく片幅	2.770	3.428	−0.658
花べん長	4.260	1.462	2.798
花べん幅	1.326	0.246	1.080

表 8d.3β　98 D.F. の併合行列 (s_{ij})

0.195340	0.092200	0.099626	0.033055
	0.121079	0.047175	0.025251
		0.125488	0.039586
			0.025106

フィッシャーの判別関数　フィッシャーは2群の母集団間の判別関数を，比

$$(\text{平均値の差})^2 \div \text{分散}$$

が最大になるような各特性の1次関数として定義した．$l_1U_1+\cdots+l_pU_p$ を1次関数とし，δ_i は2群の母集団における U_i の期待値の差とする．そうすれば，最大にすべき量は

$$(\sum l_i\delta_i)^2 \div \sum\sum l_il_j\sigma_{ij} \qquad (8d.3.3)$$

である．l_i について微分して0とおくと，方程式

$$l_1\sigma_{i1}+\cdots+l_p\sigma_{ip} = c\delta_i, \qquad i=1,\cdots,p \qquad (8d.3.4)$$

が得られる，ただし，c は定数であり，1とおいてよい．何となれば，一意に定まるのは l_i の比のみだからである．

この係数の推定値は，方程式

$$l_1s_{i1}+\cdots+l_ps_{ip} = d_i, \qquad i=1,\cdots,p \qquad (8d.3.5)$$

を解いて得られる．*Iris Setosa* と *Iris Versicolor* の問題では，表 8d.3α の平均値の差と，表 8d.3β の (s_{ij}) を用いて，解は

$$l_1=-3.0524,\quad l_2=-18.0231,\quad l_3=21.7656,\quad l_4=30.8446$$

となる．したがって，推定判別関数は

$$-3.0524U_1 - 18.0231U_2 + 21.7656U_3 + 30.8446U_4$$

となる．

D_p^2 の計算　容易にわかるように

$$D_p^2 = \sum\sum s^{ij}d_id_j = \sum l_id_i \qquad (8d.3.6)$$

であるので，D_p^2 は方程式 (8d.3.5) を解いて得られる判別関数の係数が得られれば，容易に計算できる．もし判別関数が必要でなければ，2次形式 (8d.3.6) は (8d.1.4) のように計算すればよい．今の例では，式 (8d.3.6) を用いて

$$D_4^2 = (-3.0524)(0.930)+\cdots+(30.8446)(1.080) = 103.2328$$

を得る．仮説 $\delta_i=0,\ i=1,\cdots,4$ を検定するために，(8d.3.2) の F は

$$F = \frac{95}{4}\cdot\frac{50\times50}{100\times98}\cdot103.2328 = 625.4538$$

となり，これは，4と95D.F. の分散比としては，1%水準で有意である．

付加情報の検定　D_q^2 と書いて，特性の q 個の部分集合に基づく D^2 とする．残りの $p-q$ 個の特性が判別のための付加情報をもたらさないという仮説を検定するためには，分散比検定は (8c.4.11) より，$p-q$ および (N_1+N_2-p-1)D.F. の

$$F = \frac{N_1+N_2-p-1}{p-q}\cdot\frac{N_1N_2(D_p^2-D_q^2)}{(N_1+N_2)(N_1+N_2-2)+N_1N_2D_q^2} \qquad (8d.3.7)$$

に基づくものとなる．

Iris Versicolor と *Iris Setosa* の例で，判別のためにはがく片長と花べん長だけで十

分であるかどうかを調べよう．長さをあらわす特性だけで計算した D_2^2 は 76.7323 となり，すべての測定値による D_4^2 は 103.2328 であるので，

$$F=\frac{95}{2}\cdot\frac{50\times 50(103.2328-76.7323)}{100\times 98+50\times 50(76.7323)}=\frac{95}{2}(0.328577)=15.6074$$

が得られ，この値は 2 と 95 D. F. では有意に高いので，幅は判別に関して長さが与える以外の情報を持ち，幅の差は長さの差だけによって決るものではないことが示される．

指定された判別関数にたいする検定 (Fisher (1940))　2つの種の間の判別関数は，$D_4^2=103.2328$ の値をもつ，

$$-3.0524U_1-18.0231U_2+21.7656U_3+30.8446U_4$$

と推定された．*Versicolor* の測定値の平均は，がく片幅以外は *Setosa* の平均より大きいので，

$$y=U_1-U_2+U_3+U_4 \tag{8d.3.8}$$

という形の判別関数も考えられる．このようなとき，1次関数 (8d. 3. 8) が2つの種の間の判別に十分であるかどうかを知ることに興味をもつかもしれない．この目的のために，1次関数 y に基づく D_1^2 を計算して，検定 (8d. 3. 7) を適用する．

$$D_1^2=\frac{(\bar{y}_1-\bar{y}_2)^2}{s_{yy}}$$

ただし，$\bar{y}_1, \bar{y}_2$ は2群の標本における平均値であり，s_{yy} は y の推定分散である．さて，

$$\begin{aligned}\bar{y}_1-\bar{y}_2&=d_1-d_2+d_3+d_4\\&=0.930+0.658+2.798+1.080=5.466\end{aligned}$$

$$\begin{aligned}V(y)&=V(U_1)+V(U_2)+V(U_3)+V(U_4)-2C(U_1,U_2)+2C(U_1,U_3)\\&\quad+2C(U_1,U_4)-2C(U_2,U_3)-2C(U_2,U_4)+2C(U_3,U_4)\end{aligned}$$

であり，これより $V(y)$ の推定値が得られる．表 8d. 3β の s_{ij} の値を用いると，

$$\begin{aligned}s_{yy}&=s_{11}+s_{22}+s_{33}+s_{44}-2s_{12}+2s_{13}+2s_{14}-2s_{23}-2s_{24}+2s_{34}\\&=0.482295\end{aligned}$$

となる．D_1^2 の値は

$$\frac{(5.466)^2}{0.482295}=61.9479$$

である．指定した判別関数の妥当性を検定するための，$p-1$ および (N_1+N_2-p-1) D. F. の分散比は

$$F=\frac{N_1+N_2-p-1}{p-1}\cdot\frac{N_1N_2(D_p^2-D_1^2)}{(N_1+N_2)(N_1+N_2-2)+N_1N_2D_1^2}$$

である．D_4^2 および D_1^2 の値を代入して，

$$F=\frac{95}{3}(0.62678)=19.8481$$

が得られる．この値は 3 と 95 D. F. で 1％ 水準で有意であり，1次関数 (8d. 3. 8) は最良の判別関数ではないことを示している．

判別関数の係数にたいする検定　判別関数の係数の標準誤差が，任意の1個の係数の有意性を判断するために，評価されている．しかしながら，判別関数の係数は，ある定まった母集団の母数の推定値であるという意味では一意でないために，この接近法には若干難かしい点がある．しかし，任意の2つの係数の比は一意であり，ある指定された比の値にたいしての厳密な検定は可能である．たとえば，i 番目と j 番目の特性にたいする仮説的な比が ρ であるとすると，U_i と U_j を $U_i+\rho U_j$ によっておきかえて得られる全体で $p-1$ 個の特性

$$U_1,\cdots,U_{i-1},U_{i+1},\cdots,U_{j-1},U_{j+1},\cdots,U_p,U_i+\rho U_j \tag{8d.3.9}$$

を考えてみよう．D_{p-1}^2 は集合 (8d. 3. 9) に基づく D^2 とし，D_p^2 は p 特性よりなる全体の集合に基づくとする．ρ が真の値であるかどうかの検定のための分散比は，1 と N_1+N_2-p-1 D. F. で，

$$F=\frac{N_1+N_2-p-1}{1}\cdot\frac{N_1N_2(D_p^2-D_{p-1}^2)}{(N_1+N_2)(N_1+N_2-2)+N_1N_2D_{p-1}^2} \tag{8d.3.10}$$

である (なお，542 ページ問題 **15** を参照せよ)．

8d. 4　数個の母集団の間の平均値の差の検定

k 個の母集団より，大きさ $N_1,\cdots,N_k$ $(N=\sum N_i)$ の標本がそれぞれ得られるとし，大きさ N_r 個の r 番目の標本の，標本平均値および補正済 S. P. 行列を

$$\bar{U}_{1r},\cdots,\bar{U}_{pr};\quad (S_{ij}^{(r)}),\quad N_r-1\ \text{D. F.}$$

と書く．さらに，

$$\bar{U}_1,\cdots,\bar{U}_p;\quad (S_{ij}),\quad \sum N_r-1\ \text{D. F.}$$

を，全標本を併合したときの，総平均値と補正済 S. P. 行列とする．母集団間の積和は，平均値のかわりに合計を使って，次式により計算する．

$$B_{ij}=\sum_{r=1}^{k}N_r\bar{U}_{ir}\bar{U}_{jr}-N\bar{U}_i\bar{U}_j=\sum_{r=1}^{k}\frac{T_{ir}T_{jr}}{N_r}-\frac{T_iT_j}{N}$$

母集団内の積和は，

$$W_{ij}=\sum_{r=1}^{k}S_{ij}^{(r)}=S_{ij}-B_{ij}$$

であり，したがって，W_{ij} は個々の補正済 S. P. 行列を実際に計算しないで求めることができる．このようにして，1変量分散分析の場合と同様に，母集団間と母集団内への分散共分散分析を行うことができる．

分散共分散分析

	D.F.	S.P. 行列
群（母集団）間	$k-1$	(B_{ij})
群（母集団）内	$N-k$	(W_{ij})
全　　体	$N-1$	(S_{ij})

Λ 基準は，

$$\Lambda = |W| \div |W+B|$$

であり，**8c.5** の記号で，(8c.5.3) を適用するためには，

$$m = t - \frac{p+q+1}{2}, \qquad t = N-1, \quad q = k-1$$

$$\lambda = \frac{pq-2}{4}, \quad s = \sqrt{(p^2q^2-4)}/\sqrt{p^2+q^2-5}, \qquad r = \frac{pq}{2}$$

である．$p(k-1)$ D.F. の χ^2 として $-m\log_e\Lambda$ を用いるか，より正確には，$2r$ と $(ms-2\lambda)$ D.F. の分散比として

$$F = \frac{ms-2\lambda}{2r}\frac{1-\Lambda^{1/s}}{\Lambda^{1/s}} \tag{8d.4.1}$$

を用いることができる．

表 8d.4α に，インドのある都市の6つの学校のほとんど同じ年齢の140人の男生徒を測定して得られた，頭の長さ U_1，頭の幅 U_2，体重 U_3 の3特性についての分散共分散分析を示す．15 D.F. の χ^2 の値として，26.0559 は5％水準で有意である．分散比による近似値を使うために，$2r=15$, $s=2.76$, $ms-2\lambda=364.72$ を計算する．分散比 (8d.4.1) は1.77となり，これは D.F. 15と364.72で5％水準で有意である．群内の S.P. 行列の

表 8d.4α　分散共分散分析

	D.F.	S.P. 行列 U_1^2	U_2^2	U_3^2	U_1U_2	U_1U_3	U_2U_3
学校間	5	752.0	151.3	1,612.7	214.2	521.3	401.2
学校内	134	12,809.3	1,499.6	21,009.6	1,003.7	2,671.2	4,123.6
全　体	139	13,561.3	1,650.9	22,622.3	1,217.9	3,192.5	4,524.8

$$\Lambda = \begin{vmatrix} 12,809.3 & 1,003.7 & 2,671.2 \\ 1,003.7 & 1,499.6 & 4,123.6 \\ 2,671.2 & 4,123.6 & 21,009.6 \end{vmatrix} \div \begin{vmatrix} 13,561.3 & 1,217.9 & 3,192.5 \\ 1,217.9 & 1,650.9 & 4,524.8 \\ 3,192.5 & 4,524.8 & 22,622.3 \end{vmatrix}$$

$$-\log_e\Lambda = 0.193724, \quad m = 139 - \tfrac{1}{2}(3+5+1) = 134.5$$

$$\chi^2 = -m\log_e\Lambda = 26.0559, \quad \text{D.F.} = 15$$

自由度が大きいことを考えれば，どちらも良い近似となっている．

8d.5　頭蓋骨特性における永年変異に関するバーナードの問題

Barnard (1935) の研究した頭蓋骨特性の定向変異についての測定は，生物測定学的研究においてある種の重要性をもっている．彼女の研究には2つの問題が含まれている．

1. 頭蓋骨の7つの特性からより少数の特性を選び出すこと．これらの7特性は，4つの時代区分に属するエジプト人の頭蓋骨において時間とともに起った変化に関して，最大限に意味のある情報をもっている．
2. 測定値の1次式を決定すること．その1次式は頭蓋骨の漸進的変化に関して，個々の頭蓋骨を最も効果的に特徴づけるように選ぶ．

歯槽基部長 U_1，鼻高 U_2，最大頭幅 U_3 および前頂基部高 U_4 の4種の測定値をとりあげると，表 8d.5α に示すように，それに関係する分散共分散分析が得られる．

表 8d.5α　分散共分散分析

特　性	群間 回　帰 (1 D.F)	偏　差 (2 D.F.)	全　体 (3 D.F.)	群　内 (394 D.F.)	全　体 (397 D.F.)
U_1^2	119.9303	3.2503	123.1806	9661.9975	9785.1781
U_2^2	459.7344	26.6115	486.3459	9073.1150	9559.4609
U_3^2	39.0429	111.3686	150.4115	3938.3204	4088.7319
U_4^2	124.8741	515.8598	640.7339	8741.5088	9382.2427
U_1U_2	−234.8108	3.4352	−231.3756	445.5733	214.1977
U_1U_3	68.4282	18.8771	87.3053	1130.6239	1217.9292
U_1U_4	−122.3772	−6.3868	−128.7640	2148.5842	2019.8202
U_2U_3	−133.9752	26.4696	−107.5056	1239.2220	1131.7164
U_2U_4	239.6016	−114.2883	125.3133	2255.8127	2381.1260
U_3U_4	−69.8243	−67.7565	−137.5808	1271.0547	1133.4739
行　列	(R_{ij})	(D_{ij})	(B_{ij})	(W_{ij})	$(S_{ij}^{(0)})$

4つの時代の群間 (3 D.F.) と群内 (394 D.F.) の S.P. 行列は **8d.4** と同様にして求められる．さらに，群間の S.P. 要素を時間による1次回帰 (1 D.F.) と回帰からの偏差 (2 D.F) に分解するために，回帰分析を行う．4つの時代にたいする時間変数 t の値は $-5, -1, 1$ および 5 とする．1次回帰の計算に出てくる値は

$$\sum(t-\bar{t})^2 = 4307.6683$$

$$\sum U_1(t-\bar{t}) = 718.7628, \qquad \sum U_3(t-\bar{t}) = 410.1019$$

$$\sum U_2(t-\bar{t}) = -1407.2608, \qquad \sum U_4(t-\bar{t}) = -733.4276$$

である．回帰による積和は

$$R_{11} = (\sum U_1(t-\bar{t}))^2 \div \sum (t-\bar{t})^2 = 119.9303$$

$$R_{12} = (\sum U_1(t-\bar{t}))(\sum U_2(t-\bar{t})) \div \sum (t-\bar{t})^2 = -234.8108$$

の式から求める．以下同様．

問題 1.　特性 U_3, U_4 は U_1, U_2 と独立に時代間で有意な変動を示すか？

付加情報を調べるための検定基準は，U_1, U_2, U_3, U_4 および U_1, U_2 のそれぞれにたいするウイルクスの Λ の値である Λ_4 と Λ_2 の比である．

$$\Lambda_4 = \frac{|W_{ij}|}{|S_{ij}^{(0)}|}, \quad i, j = 1, \cdots, 4\,; \quad \Lambda_2 = \frac{|W_{ij}|}{|S_{ij}^{(0)}|}, \quad i, j = 1, 2.$$

表 8d. 5α から $S_{ij}^{(0)}$ および W_{ij} の値を代入すれば，$\Lambda_4/\Lambda_2 = 0.87806$ となり，**8c. 5** における記号で，$p=2$，$q=3$，$t=395$ より $m=392$ が与えられ，

$$\chi^2 = -392 \log_e(0.87806) = 50.98, \qquad 6 \text{ D.F.}$$

となることがわかる．この値は大きく，4 特性全体が重要であることを示している．

問題 2.　各特性の時間にたいする回帰は直線と考えることができるか？

この問題にたいする適切な検定基準は，$p=4$，$q=2$，$t=396$ で

$$\Lambda = \frac{|W|}{|D+W|} = 0.8990$$

となる．χ^2 の近似値は，8 D.F. で 41.79 と大きく，回帰の非線形性を示している．この検定は，容易にわかるように，平均値の差を時間にたいする 2 次以上の回帰で説明できるかどうかを調べるために拡張可能である．

この他の応用については，Bartlett (1947)，Riffenburg and Clunies-Ross (1960)，Williams (1952, 1955, 1959) および著者の論文，Rao (1948b, 1949b, 1954c, 1955a, 1958b, 1959a, 1961g, 1962c, 1965b, 1965d, 1966c, f, 1967c, 1972a) を参照せよ．同時に，関連のある問題の議論と応用に関しては，Roy, Gnanadesikan and Srivastava (1972) および Williams (1967) を見よ．

8e　判別分析（識別問題）

8e. 1　決定のための判別得点

統計的推測を取り扱った第 7 章の **7d. 1** から **7d. 6** までで，観察された個体が，それが所属する可能性のある所与の 1 組の母集団のなかのどれかのメンバーであることを決定する問題，すなわち，識別の問題を議論した．この問題に満足な解を与えるためには，次のことを知らねばならない．

1. k 個の異なる母集団に属する個体についての，所与の一組の測定値 $\mathbf{U}$ にたいする確率密度，$P_1(\mathbf{U}), \cdots, P_k(\mathbf{U})$.
2. 識別すべき個体が取り出された複合集団における，k 個の母集団の個体の相対度数にあたるところの，母集団の先験的確率 $\pi_1, \cdots, \pi_k$.
3. 第 i 母集団の個体を誤って第 j 母集団のメンバーと識別するときの損失をあらわす値 r_{ij} の規定，すなわち損失関数の決定．

このとき，測定値 $\mathbf{U}$ をもつ個体が得られると，その第 i 母集団への判別得点は

$$S_i = -[\pi_1 P_1(\mathbf{U}) r_{1i} + \cdots + \pi_k P_k(\mathbf{U}) r_{ki}], \qquad i = 1, \cdots, k \qquad (8e.1.1)$$

と計算され（(7d. 3. 8) 参照），その個体は，この判別得点が最高になる母集団にわりつけられる．この規則は，期待損失（長い眼で見た，**7d. 3** 参照）を最小にすることがわかる．

この決定規則 (8e. 1. 1) は，職業補導や選抜問題のような日常の仕事で，個体を識別しなければならないときには，理想的である．そのような場合には，正しい識別は何であるかが，後の時点でわかるようになるかもしれないし，データを集積すると，密度 $P_i(\mathbf{U})$ および先験的確率 π_i の十分良い推定値が得られるようになるであろうから，規則(8e. 1. 1)の適用が可能になる．

多くの実際問題では，間違った識別による損失を評価することは難かしく，その場合には，間違った識別をする度数（多くの試行のなかでの）を最小にするという基準が有用であるかもしれない．そのような場合に最適な規則は，測定値 $\mathbf{U}$ をもつ個体を，その事後確率が最大値をとる母集団にわりつけることである．これを，別の言葉で言えば，第 i 母集団への判別得点を

$$S_i = \pi_i P_i(\mathbf{U}) \qquad (8e.1.2)$$

とする〔そして，これを最大にする母集団を選ぶ〕ことである．

$\mathbf{U}$ の分布が各母集団で p 変量正規分布である場合を考える．$P_i(\mathbf{U})$ を正規確率密度，

$$(2\pi)^{-p/2}|\Sigma_i|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{U}-\boldsymbol{\mu}_i)'\Sigma_i^{-1}(\mathbf{U}-\boldsymbol{\mu}_i)\right], \qquad i = 1, \cdots, k \qquad (8e.1.3)$$

にとる．すなわち第 i 母集団では平均 $\boldsymbol{\mu}_i$，分散共分散行列 Σ_i をもつとする．$\pi_i P_i(\mathbf{U})$ の対数をとり，すべての i に共通な因子 $(2\pi)^{-p/2}$ を除くと，第 i 母集団にたいする (8e. 1. 2) と等価な判別得点は，

$$S_i = -\frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(\mathbf{U}-\boldsymbol{\mu}_i)'\Sigma_i^{-1}(\mathbf{U}-\boldsymbol{\mu}_i) + \log \pi_i \qquad (8e.1.4)$$

となり，これも第 i 母集団の平均 $\boldsymbol{\mu}_i$ と分散共分散行列 Σ_i を含む．関数 (8e. 1. 4) は，$\mathbf{U}$ の 2 次式であるから，2 次判別得点と呼んでもよいであろう．各個体は，その 2 次判別得点が最大値をとる母集団にわりつけられる．

k 個の母集団の分散共分散行列が異ならないときには，

$$-\frac{1}{2}\log|\Sigma| - \frac{1}{2}\mathbf{U}'\Sigma^{-1}\mathbf{U} \tag{8e.1.5}$$

という項は，すべての S_i に共通になる．(8e.1.4) から (8e.1.5) の項を引くと，第 i 母集団にたいする等価な判別得点は，

$$S_i = (\mu_i'\Sigma^{-1})\mathbf{U} - \frac{1}{2}\mu_i'\Sigma^{-1}\mu_i + \log\pi_i \tag{8e.1.6}$$

と書け，$\mathbf{U}$ の1次式となるから，1次判別得点とよんでよいであろう．

母集団が2つだけのときには，比較はただ1つとなり，差 $S_1 - S_2$ を計算するだけで決定を下すことができる．この値は $L(\mathbf{U}) - c$ に等しく，

$$L(\mathbf{U}) = (\mu_1' - \mu_2')\Sigma^{-1}\mathbf{U}$$

$$c = \frac{1}{2}(\mu_1'\Sigma^{-1}\mu_1 - \mu_2'\Sigma^{-1}\mu_2) + \log\pi_2 - \log\pi_1 \tag{8e.1.7}$$

と書ける．関数 (8e.1.7) は，**8d.3** で導いたフィッシャーの線形判別関数である．このとき，決定規則は

$$L(\mathbf{U}) \geqslant c \text{ なら，第1母集団にわりつけよ}$$
$$L(\mathbf{U}) \leqslant c \text{ なら，第2母集団にわりつけよ} \tag{8e.1.8}$$

という形式であらわされる．しかしながら，(8e.1.6) のような判別得点であらわす方法は，2つの母集団という特殊な場合をも，識別規則の一般的な枠組みのなかに取り込んでいる．2つの母集団の場合でも，その各々にたいする判別得点を計算すると，情報がより増えるから，実際にはそれだけ望ましい．

適用例として，精神医学的検査を受けた個人の神経症状を識別する問題を考えよう．表8e.1α に，3種の測定値 A, B, C についての過去のデータから推定された平均値と分散共分散行列を示す．

任意の群にたいする1次判別得点は，

$$l_1A + l_2B + l_3C - \frac{1}{2}(l_1m_1 + l_2m_2 + l_3m_3) + \log\pi$$

となる，ただし，A, B, C は観察された測定値であり，m_1, m_2, m_3 は平均値，π は事前確率である．係数 l_1, l_2, l_3 は，連立方程式

$$l_1 = \sigma^{11}m_1 + \sigma^{12}m_2 + \sigma^{13}m_3$$
$$l_2 = \sigma^{21}m_1 + \sigma^{22}m_2 + \sigma^{23}m_3$$
$$l_3 = \sigma^{31}m_1 + \sigma^{32}m_2 + \sigma^{33}m_3$$

から計算される．神経症状群ごとの判別得点を計算するのに必要な l_1, l_2, l_3 やその他の項は，表8e.1β に示してある．

表 8e.1α　神経症状群ごとの平均得点 (Rao and Slater (1949))

群	標本の大きさ	平均得点 A	平均得点 B	平均得点 C
不安状態	114	2.9298	1.1667	0.7281
ヒステリー	33	3.0303	1.2424	0.5455
精神病	32	3.8125	1.8438	0.8125
強迫観念	17	4.7059	1.5882	1.1176
人格変容	5	1.4000	0.2000	0.0000
正常	55	0.6000	0.1455	0.2182

群内分散共分散行列 (σ_{ij})	A	B	C	左の逆行列 (σ^{ij})	A	B	C
A	2.3008	0.2516	0.4742	A	0.5432	−0.2002	−0.4208
B	0.2516	0.6075	0.0358	B	−0.2002	1.7258	0.0558
C	0.4742	0.0358	0.5951	C	−0.4208	0.0558	2.0213

表8e.1β の最後（右）の列を計算するとき，事前確率は観察された標本の大きさに比例するとした．もっともそれは真の事前確率の推定値ではないかもしれない．適切な推定値は，過去に病院に来た種々の精神病症状の人々の相対度数である．A, B, C が測定された人にたいしては，6つの得点

$$S_i = l_1A + l_2B + l_3C + (\log\pi_i - \frac{1}{2}\sum l_jm_j), \qquad i = 1, \cdots, 6$$

を計算し，その得点が最高になる群にその人をわりつける．

表 8e.1β　判別得点を計算するための係数やその他の項

群	l_1	l_2	l_3	$\frac{1}{2}\sum l_jm_j$	$\log\pi - \frac{1}{2}\sum l_jm_j$
不安状態	1.0515	1.4676	0.2974	2.5047	−3.3137
ヒステリー	1.1678	1.5679	−0.1081	2.7139	−4.7626
精神病	1.3599	2.4641	0.1336	4.9182	−6.9977
強迫観念	1.7680	1.8611	0.3573	5.8375	−8.5495
人格変容	0.7204	0.0649	−0.5780	0.5107	−4.4465
正常	0.2050	0.1431	0.1947	0.0931	−1.6311

8e.2　研究における判別分析

8e.1 の手順は，長い目で見て損失（または，誤った識別の度数）を最小にすることが目的であるような場合の日常の決定においては，適用可能であり，かつ適切である．ここで

は，アフリカで発見された化石オーストラロピテクス・アフリカヌス (Australopithicus africanus) は，類人猿 (anthropoid) より人類 (hominid) に近いか否かを調べるような問題を考えよう．その類似性を識別するためには，アフリカヌスについてのいくつかの計測値を，人類と類人猿という既知の化石群の計測値と比較しなければならないことは明らかである．

このような場合に，観察された化石が既知の化石群のどれか1つメンバーであるということを，問題の1つのデータとして認めることは現実的でない．統計的解析においては，その化石が，存在をこれから確認しなければならないような未知の群に属するかも知れないという可能性を考慮に入れなければならない．また，既知の各群における計測値の平均と分散共分散行列の大まかな推定値を得ることは可能であるかも知れないが，その事前確率を求める方法はなにもないように思われる，異なる群から入手できる化石の数は，化石の母集団全体のなかで，それらの群の相対的な個数のよい指標であるとはかぎらない．さらに，まだ発見されない群の事前確率を知る可能性はまったくない．最後に，損失関数の概念はこのような事情のもとでは意味をもたない．したがって，**8e.1** の方法は，当面の問題解決のためには，適切ではないし，可能でもない．

そこで，2つの既知の群のどちらか，または，可能性として第3の未知の群への，個体の所属関係を調べる一般的な接近方法を論じよう．この目的のために，次を証明する．

(i) 分散共分散行列の等しい2つの p 変量正規母集団 $N_p(\mu_1, \Sigma)$ と $N_p(\mu_2, \Sigma)$ があるとき，(8e.1.7) で定義した線形判別関数（これは，定数項を除いて，確率密度の比の対数に等しい）は，

$$L(\mathbf{U}) = (\mu_2 - \mu_1)'\Sigma^{-1}\mathbf{U} \tag{8e.2.1}$$

である．いま，μ_1, μ_2，および Σ はすべて既知とする．このとき，統計量 $L(\mathbf{U})$ は，α と β が $\alpha+\beta=1$ の条件の下で変わる正規確率密度 $N_p(\alpha\mu_1+\beta\mu_2, \Sigma)$ の族において十分統計量である．いいかえると $L(\mathbf{U})$ を与えたときの $\mathbf{U}$ の条件づき分布は，$\alpha+\beta=1$ なる $\alpha\mu_1+\beta\mu_2$ のすべての値にたいして同じである．

$N_p(\alpha\mu_1+\beta\mu_2, \Sigma)$ の密度

$$(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{U}-\mu_1-\beta\delta)'\Sigma^{-1}(\mathbf{U}-\mu_1-\beta\delta)\right] \tag{8e.2.2}$$

を考える．ただし，$\alpha = 1-\beta$，$\delta = \mu_2-\mu_1$ である．指数のベキの式は ($-\frac{1}{2}$ を除くと)，

$$\begin{aligned}(\mathbf{U}-\mu_1)'\Sigma^{-1}(\mathbf{U}-\mu_1) &- 2\beta\delta'\Sigma^{-1}\mathbf{U} + 2\beta\delta'\Sigma^{-1}\mu_1 + \beta^2\delta'\Sigma^{-1}\delta \\ = H(\mathbf{U}) \qquad &- 2\beta L(\mathbf{U}) \quad + G(\beta)\end{aligned} \tag{8e.2.3}$$

となる．ただし，$H(\mathbf{U})$ と $L(\mathbf{U})$ は β を含まず，$G(\beta)$ は $\mathbf{U}$ を含まない．このとき確率密度 (8e.2.2) は，(定数項を除くと) 次の積の形に書ける．

$$\exp\left[-\frac{1}{2}H(\mathbf{U})\right]\times\exp\left\{\frac{1}{2}[2\beta L(\mathbf{U})-G(\beta)]\right\}$$

ただし，この第1項は β に無関係，第2項は $L(\mathbf{U})$ と β だけを含んでいる．したがって，$L(\mathbf{U})$ は β にたいして十分統計量である．すなわち，$L(\mathbf{U})$ を与えたときの $\mathbf{U}$ の条件つき分布は $\alpha+\beta=1$ の下で，$\alpha\mu_1+\beta\mu_2$ のすべての値にたいして同じである．

(i) の結果は，Smith (1947) の定理の一般化である．彼は，$\beta=0$ と $\beta=1$ という特殊の場合についてのみ，$L(\mathbf{U})$ の十分性を証明した．

いま，観察された個体が，2つの所与の母集団 A, B のどちらか，または，あるまだ特定されない母集団に属するか否かを調べる問題があったとしよう．(i) の結果は，A と B の確率密度から構成される判別関数が，A と B に関してのみならず，A と B の平均値を結ぶ直線上にその平均値をもつより広い母集団の組に関しても，個体の所属関係を調べるのに十分であることを示す．それゆえ，個体の所属関係について推測するのに，A と B を参照して構成された判別関数を用いるためには，A と B の平均ベクトルを結ぶ直線によって定義されるより広い母集団の組にその個体が属するか否かをまず検定する必要がある．そのような検定方法は，次の (ii) で考えられる．

(ii) 計測値 $\mathbf{U}$ をもつ個体が $N_p(\alpha\mu_1+\beta\mu_2, \Sigma)$ から来たという帰無仮説を H_0 としよう．ただし β は $\alpha+\beta=1$ の下で任意で，μ_1, μ_2, Σ は既知であるか，または，大標本に基づく推定値をもつとする．H_0 を検定するための検定基準

$$(\mathbf{U}-\mu_1)'\Sigma^{-1}(\mathbf{U}-\mu_1) - \frac{[(\mathbf{U}-\mu_1)'\Sigma^{-1}(\mu_2-\mu_1)]^2}{(\mu_2-\mu_1)'\Sigma^{-1}(\mu_2-\mu_1)} \tag{8e.2.4}$$

は，$(p-1)$ D.F. の χ^2 分布に従う．

読者は，2次形式についての諸結果を用いて，(8e.2.4) の分布を導くのに，なんの困難も感じないであろう．第1項は $\chi^2(p)$，第2項は $\chi^2(1)$ に従って分布し，その差は正であることに注目せよ．それから [(iv), **3b.4**] を適用せよ．

この検定 (8e.2.4) をアフリカヌスに適用し，2つの群としてチンパンジーと人間の化石を選び，特性として歯の長さと幅をとると，χ^2 の値はほぼ5ぐらいとなり，この値は $(p-1=1)$ D.F. にたいして大きいことがわかる．それゆえ，アフリカヌスは，チンパンジーと人間の化石の平均値に単純に結びつけられないような平均値をもつ群に属するのであろう．

検定 (8e.2.4) の結果が有意でなかったとすると，判別関数に基づくさらに進んだ解析が必要になる．その群の1つ（チンパンジー）への所属は，統計量

$$\frac{[(\mu_2-\mu_1)'\Sigma^{-1}(\mathbf{U}-\mu_1)]^2}{(\mu_2-\mu_1)'\Sigma^{-1}(\mu_2-\mu_1)} \sim \chi^2(1)$$

によって，また第2の群（人間）への所属は，

$$\frac{[(\mu_2-\mu_1)'\Sigma^{-1}(\mathbf{U}-\mu_2)]^2}{(\mu_2-\mu_1)'\Sigma^{-1}(\mu_2-\mu_1)} \sim \chi^2(1)$$

によって吟味される．もしこの両方が1 D. F. の χ^2 として有意に大きいならば，その個体は，第3の群に所属する可能性が再び生ずる．ただし，その平均ベクトルは，チンパンジーと人間の化石の平均ベクトルを結ぶ直線上にある．もし一方の χ^2 が小さくて他方が大きければ，その個体は群のどちらかに所属することが示される．また，χ^2 が両方とも小さくなる可能性もあって，その場合には，個体がどちらの群に類似しているとはっきり言えないことになる．

アフリカヌスの例では，チンパンジーにたいする χ^2 は 3.80 で，5% の有意点に近く，人間にたいする χ^2 は 0.03 で非常に小さかった．しかしながらこれらの検定は有効ではない（誤解をうけやすい）．なぜなら，この判別関数はアフリカヌスの位置を，ここで取り上げた2群に基づいて推測するためには適当ではないからである．事前の検定 (8e. 2. 4)をせずに，判別関数を無批判に用いると，アフリカヌスは人間の群に非常に近いという誤った結論に導かれる．そうではなくて，この化石は，第3の未知の群に属するであろうことが ((8e. 2. 4) の検定を使って) 証明されたのである．

上に述べた解析は，判別する群の数が多い場合にも拡張できる (Rao (1962c) 参照)．アフリカヌスに関してさらに詳細に，かつこれ以前の研究についての文献を知りたい読者は Ashton, Healy and Lipton (1957) を参照せよ．

対立仮説の母数が未知のときは，問題は複雑になるけれども，推定することができる (John (1961), Rao (1954a) および Wald (1944) 参照).

8e. 3 複合仮説間の判別

$\mathbf{U}$ を p 変量確率変数とし，$P(\cdot|\theta)$ を母数 $\theta\in\Theta$ によって定まる $\mathbf{U}$ の確率密度とする．H_1 を仮説 $\theta\in\Theta_1$ とし，H_2 を $\theta\in\Theta_2$ とする．ただし，Θ_1 と Θ_2 は2つの互いに素である Θ の部分集合である．問題は，$\mathbf{U}$ の観測値に基づいて，H_1 と H_2 のうちの1つを選択することである．

この問題は，確率密度が

$$P(s|\theta)=\begin{cases}P_1(s) & \text{任意の } \theta\in\Theta_1 \text{ にたいして}\\ P_2(s) & \text{任意の } \theta\in\Theta_2 \text{ にたいして}\end{cases} \qquad (8e.\,3.\,1)$$

であるような補助統計量 S が存在すれば，きれいな解が得られる．$P_1(\cdot)$ と $P_2(\cdot)$ が異なり，かつ θ と独立であるときは，H_1 と H_2 のいずれかを選ぶための判別関数は，尤度比

$$\log\frac{P_1(s)}{P_2(s)} \qquad (8e.\,3.\,2)$$

となり，これは Θ_1 と Θ_2 における個々の仮説にはよらない．いくつかの例を考えよう．

例 1. $\mathbf{U}$ を次のような p 次元ベクトルの正規確率変数とする．

$$\begin{aligned} &E(\mathbf{U}|\theta_1, H_1)=\alpha_1+\mathbf{B}'\theta_1, && D(\mathbf{U}|\theta_1, H_1)=\Sigma,\\ &E(\mathbf{U}|\theta_2, H_2)=\alpha_2+\mathbf{B}'\theta_2, && D(\mathbf{U}|\theta_2, H_2)=\Sigma. \end{aligned} \qquad (8e.\,3.\,3)$$

これらの母数 θ_1 と θ_2 は未知であるが，複合仮説 H_1 と H_2 によって定まる所与の集合に属するとする．問題は観測値 $\mathbf{U}$ に基づいて，H_1 と H_2 の何れが真であるかを決定することである．このような問題は Burnaby (1964) が2群の化石を判別したときに起こった．各化石群は異なる生長期にある個体で構成される部分母集団の混合したものであった．問題は，化石の年齢が未知のとき，2群のいずれに属する化石であるかを識別することである．バーナビーの場合には

$$E(\mathbf{U}|t, H_1)=\alpha_1+t\beta, \qquad E(\mathbf{U}|t, H_2)=\alpha_2+t\beta \qquad (8e.\,3.\,4)$$

となる．ただし，t は年齢を示し，β は各種の形質の生長率を示すベクトルである．(8e. .4) がその特殊な場合である一般模型 (8e. 3. 3) を考えよう．$k=R(\mathbf{B})$ として，$\mathbf{BC}'=0$ となる階数 $p-k$ の $(p-k)\times p$ 行列 $\mathbf{C}$ を考える．そうすれば

$$\begin{aligned} &E(\mathbf{CU}|H_1)=\mathbf{C}\alpha_1, && D(\mathbf{CU}|H_1)=\mathbf{C}\Sigma\mathbf{C}',\\ &E(\mathbf{CU}|H_2)=\mathbf{C}\alpha_2, && D(\mathbf{CU}|H_2)=\mathbf{C}\Sigma\mathbf{C}', \end{aligned} \qquad (8e.\,3.\,5)$$

は θ_1, θ_2 に無関係で，$\mathbf{CU}$ は (8e. 3. 1) 型の補助変数の1つになることがわかる．正規確率密度を用いて，$\mathbf{CU}$ に基づく判別関数は

$$(\mathbf{C}\alpha_1-\mathbf{C}\alpha_2)'(\mathbf{C}\Sigma\mathbf{C}')^{-1}\mathbf{CU}=(\alpha_1-\alpha_2)'\mathbf{C}'(\mathbf{C}\Sigma\mathbf{C}')^{-1}\mathbf{CU} \qquad (8e.\,3.\,6)$$

となる．第1章末の問題 33 (72頁) の恒等式を用いて，

$$\mathbf{C}'(\mathbf{C}\Sigma\mathbf{C}')^{-1}\mathbf{C}=\Sigma^{-1}-\Sigma^{-1}\mathbf{B}'(\mathbf{B}\Sigma^{-1}\mathbf{B}')^{-1}\mathbf{B}\Sigma^{-1}. \qquad (8e.\,3.\,7)$$

(8e. 3. 6) は

$$(\alpha_1-\alpha_2)'(\Sigma^{-1}-\Sigma^{-1}\mathbf{B}'(\mathbf{B}\Sigma^{-1}\mathbf{B}')^{-1}\mathbf{B}\Sigma^{-1})\mathbf{U} \qquad (8e.\,3.\,8)$$

と変形できる．この結果 (8e. 3. 8) は既知の量 $\alpha_1, \alpha_2, \mathbf{B}$ および Σ だけであらわされる．

例 2. こんどは別の複合仮説を考えよう．

$$\begin{aligned} &E(\mathbf{U}|\theta_1, H_1)=\alpha_1+\mathbf{B}'\theta_1, && D(\mathbf{U}|\theta_1, H_1)=\Sigma_1,\\ &E(\mathbf{U}|\theta_2, H_2)=\alpha_2+\mathbf{B}'\theta_2, && D(\mathbf{U}|\theta_2, H_2)=\Sigma_2. \end{aligned} \qquad (8e.\,3.\,9)$$

ただし，θ_1, θ_2 は未知とする．

例1で定義したと同じ統計量 $\mathbf{CU}$ は (8e. 3. 9) における2つの仮説の下でも補助統計量である．H_1 と H_2 のもとでの分布は

$$\begin{aligned} &E(\mathbf{CU}|H_1)=\mathbf{C}\alpha_1, && D(\mathbf{CU}|H_1)=\mathbf{C}\Sigma_1\mathbf{C}',\\ &E(\mathbf{CU}|H_2)=\mathbf{C}\alpha_2, && D(\mathbf{CU}|H_2)=\mathbf{C}\Sigma_2\mathbf{C}' \end{aligned} \qquad (8e.\,3.\,10)$$

によって定まる．そうすれば対数尤度比は $\mathbf{U}$ の2次形式

$$\mathbf{U}'\mathbf{C}'[(\mathbf{C}\Sigma_1\mathbf{C}')^{-1}-(\mathbf{C}\Sigma_2\mathbf{C}')^{-1}]\mathbf{C}\mathbf{U}-2[\alpha_1'\mathbf{C}'(\mathbf{C}\Sigma_1\mathbf{C}')^{-1}-\alpha_2'\mathbf{C}'(\mathbf{C}\Sigma_2\mathbf{C}')^{-1}]\mathbf{C}\mathbf{U} \tag{8e.3.11}$$

になる．各 Σ_i にたいして恒等式(8e. 3. 7)を用いて，(8e. 3. 11) は α_i, Σ_i および $\mathbf{B}$ のみによって表現できる．$\mathbf{U}$ の分布に関してはなにも仮定しないで，判別関数 (8e. 3. 8) と (8e. 3. 11) の性質を調べよう．そうするとき，$\mathbf{U}$ に関する過去のデータは H_1 と H_2 における母集団の混合についてだけ得られるという状況をも考察しよう．そのようなときには，$\alpha_1, \alpha_2, \Sigma_1, \Sigma_2$ そのものではなく，次の母数のみが既知であるか，または推定できる．

$$E(\mathbf{U}|H_1)=\alpha_1+\mathbf{B}'\bar{\theta}_1,\qquad D(\mathbf{U}|H_1)=\Sigma_1+\mathbf{B}'\mathbf{D}_1\mathbf{B}$$
$$E(\mathbf{U}|H_2)=\alpha_2+\mathbf{B}'\bar{\theta}_2,\qquad D(\mathbf{U}|H_2)=\Sigma_2+\mathbf{B}'\mathbf{D}_2\mathbf{B}$$

ここで，，$\bar{\theta}_1, \bar{\theta}_2$ は θ_1 と θ_2 の未知の平均値，$\mathbf{D}_1, \mathbf{D}_2$ は H_1 と H_2 における混合に依存する未知の行列である．

$\delta=\alpha_1-\alpha_2+\mathbf{B}'(\bar{\theta}_1-\bar{\theta}_2)$，$\Lambda_1=\Sigma_1+\mathbf{B}'\mathbf{D}_1\mathbf{B}$ および $\Lambda_2=\Sigma_2+\mathbf{B}'\mathbf{D}_2\mathbf{B}$ とする．もし，$\alpha_1-\alpha_2$ のかわりに δ を，Σ のかわりに $\Lambda=\Sigma+2^{-1}\mathbf{B}'(\mathbf{D}_1+\mathbf{D}_2)\mathbf{B}$ を用い，Σ_1, Σ_2 のかわりに $\Lambda_1=\Sigma_1+\mathbf{B}'\mathbf{D}_1\mathbf{B}$，$\Lambda_2=\Sigma_2+\mathbf{B}'\mathbf{D}_2\mathbf{B}$ を用いると，$\bar{\theta}_1, \bar{\theta}_2, \mathbf{D}_1, \mathbf{D}_2$ が何であっても，判別関数 (8e. 3. 8) および (8e. 3. 11) はそのままの形で残る．これは極めて都合のよいことである．

まず，1次判別関数 (8e. 3. 8) を特徴づける．

(i)　$\mathbf{U}$ を次のようなベクトル確率変数とする．

$$E(\mathbf{U}|H_1)=\alpha_1+\mathbf{B}'\bar{\theta}_1,\qquad D(\mathbf{U}|H_1)=\Sigma+\mathbf{B}'\mathbf{D}_1\mathbf{B},$$
$$E(\mathbf{U}|H_2)=\alpha_2+\mathbf{B}'\bar{\theta}_2,\qquad D(\mathbf{U}|H_2)=\Sigma+\mathbf{B}'\mathbf{D}_2\mathbf{B}. \tag{8e.3.12}$$

そうすれば，

$$\sup_{\mathbf{BL}=0}\frac{[E(\mathbf{L}'\mathbf{U}|H_1)-E(\mathbf{L}'\mathbf{U}|H_2)]^2}{2^{-1}[V(\mathbf{L}'\mathbf{U}|H_1)+V(\mathbf{L}'\mathbf{U}|H_2)]} \tag{8e.3.13}$$

は次のとき得られる．

$$\mathbf{L}_*=[\Sigma^{-1}-\Sigma^{-1}\mathbf{B}'(\mathbf{B}\Sigma^{-1}\mathbf{B}')^{-1}\mathbf{B}\Sigma^{-1}](\alpha_1-\alpha_2) \tag{8e.3.14}$$

条件 $\mathbf{BL}=0$ のもとで，式 (8e. 3. 13) は

$$\sup_{\mathbf{BL}=0}\frac{[\mathbf{L}'(\alpha_1-\alpha_2)]^2}{\mathbf{L}'\Sigma\mathbf{L}} \tag{8e.3.15}$$

と変形され，この結果は (1c. 6. 3) (48ページ) を適用して導かれる．

判別関数 (8e. 3. 8) は，$\mathbf{L}_*$ が (8e. 3. 14) で定義されていれば，$\mathbf{L}_*'\mathbf{U}$ であることに注意せよ．2つの単純仮説にたいする1次判別は $\mathbf{BL}=0$ の制限なしに (8e. 3. 13) で上限をとることにより得られることを思い出すべきである．$\mathbf{L}_*$ は $\alpha_1-\alpha_2$ のかわりに δ を，Σ のかわりに $(\Lambda_1+\Lambda_2)/2$ を用いても同じである．

(ii)　2次判別関数 (8e. 3. 11) は，$\alpha_1-\alpha_2$ のかわりに δ を用い，Σ_1, Σ_2 のかわりに Λ_1, Λ_2 を用いても変化しない．〔α_1, α_2 のかわりに $\alpha_1+\mathbf{B}'\bar{\theta}_1$，$\alpha_2+\mathbf{B}'\bar{\theta}_2$ を用いる．〕

この結果は，直接験算によって証明できる．

8f　変量の組のあいだの関係

8f.1　正準相関

(4g. 1. 11) で定義した重相関係数は，1つの変数と1組の他の変数との関連の程度を測るものであった．それは，実は1つの変数と他の変数の1次関数とのあいだの最大の相関であり，また1つの変数と他の変数によるその最小平均2乗誤差予測子との相関でもあることが証明された [(iv), 4g. 1]．Hotelling (1935, 1936) はこの概念を一般化して，2組の変数のあいだの関連性をしらべた．

$\mathbf{U}_1'=(U_1,\cdots,U_r)$，$\mathbf{U}_2'=(U_{r+1},\cdots,U_p)$ を2組の確率変数とし，

$$D(\mathbf{U}_1)=\Sigma_{11},\qquad \mathrm{cov}(\mathbf{U}_1,\mathbf{U}_2)=\Sigma_{12},\qquad D(\mathbf{U}_2)=\Sigma_{22} \tag{8f.1.1}$$

とおく．分散1をもつ2つの1次関数 $\mathbf{L}'\mathbf{U}_1$ と $\mathbf{M}'\mathbf{U}_2$ を考え，$\mathbf{L}$ と $\mathbf{M}$ は，$\mathbf{L}'\mathbf{U}_1$ と $\mathbf{M}'\mathbf{U}_2$ との相関が最大になるように選ぶ．すると，この問題は，$\mathbf{L}'\Sigma_{11}\mathbf{L}=1=\mathbf{M}'\Sigma_{22}\mathbf{M}$ という条件の下で，$\mathbf{L}'\Sigma_{12}\mathbf{M}$ を最大にする問題となる．ラグランジュの乗数を導入すると，微分すべき式は

$$\mathbf{L}'\Sigma_{12}\mathbf{M}-\frac{\lambda_1}{2}\mathbf{L}'\Sigma_{11}\mathbf{L}-\frac{\lambda_2}{2}\mathbf{M}'\Sigma_{22}\mathbf{M}$$

となり，これから，次式を得る．

$$\Sigma_{12}\mathbf{M}-\lambda_1\Sigma_{11}\mathbf{L}=0,\qquad -\lambda_2\Sigma_{22}\mathbf{M}+\Sigma_{21}\mathbf{L}=0 \tag{8f.1.2}$$

この第1式から $\mathbf{L}'\Sigma_{12}\mathbf{M}=\lambda_1$，第2式から $\lambda_2=\mathbf{M}'\Sigma_{21}\mathbf{L}$ を得，$\lambda_1=\lambda_2=\rho$ (とおく) がわかる．(8f. 1. 2) の第1式に $\Sigma_{21}\Sigma_{11}^{-1}$ をかけ，それに第2式の ρ 倍を加えると

$$(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}-\rho^2\Sigma_{22})\mathbf{M}=0 \tag{8f.1.3}$$

を得る．それゆえ，ρ^2 と $\mathbf{M}$ は，固有方程式

$$|\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}-\rho^2\Sigma_{22}|=0 \tag{8f.1.4}$$

の固有値・固有ベクトルである．$\rho_1^2,\cdots,\rho_s^2$ $(s=p-r)$ を固有値，$\mathbf{M}_1,\cdots,\mathbf{M}_s$ を対応する固有ベクトルとする．さらに，$\mathscr{M}=(\mathbf{M}_1\vdots\cdots\vdots\mathbf{M}_s)$ とおくと，次式を得る．

$$\mathscr{M}'\Sigma_{22}\mathscr{M}=\mathbf{I},\qquad \Sigma_{22}=\mathscr{M}'^{-1}\mathscr{M}^{-1},$$
$$\mathscr{M}'\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\mathscr{M}=\mathbf{R}_2,\qquad \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}=\mathscr{M}'^{-1}\mathbf{R}_2\mathscr{M}^{-1}. \tag{8f.1.5}$$

ただし，$\mathbf{R}_2$ は固有値 $\rho_1^2,\cdots,\rho_s^2$ を対角要素とする対角行列である (2つの2次形式の正準

化に関する [(ii), **1c. 3**] 参照).

同様にして，固有方程式

$$|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}-\rho^2\Sigma_{11}|=0 \qquad (8f.1.6)$$

を得，その固有値を $\rho_1^2,\cdots,\rho_r^2$，固有ベクトルを $\mathbf{L}_1,\cdots,\mathbf{L}_r$ とする．(8f. 1. 4) と (8f. 1. 6) の非零根は同じであるから，それらの根をあらわすのに同じ記号を用いることができる．しかしながら，零の根の重複度は2つの場合で異なる．$\mathscr{L}$ と $\mathbf{R}_1$ を (8f. 1. 5) の $\mathscr{M}$ と $\mathbf{R}_2$ に対応するものとすると，

$$\mathscr{L}'\Sigma_{11}\mathscr{L}=\mathbf{I}, \qquad \Sigma_{11}=\mathscr{L}'^{-1}\mathscr{L}^{-1}$$
$$\mathscr{L}'\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\mathscr{L}=\mathbf{R}_1, \qquad \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}=\mathscr{L}'^{-1}\mathbf{R}_1\mathscr{L}^{-1} \qquad (8f.1.7)$$

を得る．このとき，非零根 $\rho_1\geqslant\rho_2\geqslant\cdots$ を正準相関と呼び，1次関数

$$\mathbf{L}_1'\mathbf{U}_1,\cdots,\mathbf{L}_r'\mathbf{U}_1 \quad \text{および} \quad \mathbf{M}_1'\mathbf{U}_2,\cdots,\mathbf{M}_s'\mathbf{U}_2$$

を正準変数と呼ぶ．

8f. 2　正準変数のもつ性質

(8f. 1. 4) と (8f. 1. 6) から求められた $\mathbf{L}_i$ と $\mathbf{M}_i$ は，ρ_i^2 の正の平方根を ρ_i とするとき，$\pm\rho_i$ を代入した (8f. 1. 2) を満足する．しかしながら，必要なら $\mathbf{L}_i$ と $\mathbf{M}_i$ の符号を変えることによって，$\mathbf{L}_i,\mathbf{M}_i,\rho_i$ を (8f. 1. 2) の解とすることができる．便宜上，正準相関はすべて正であるとする．すると，次の諸結果を得る．

(i)　(8f. 1. 4) および (8f. 1. 6) の非零根の数は Σ_{12} の階数に等しい．

(ii)　(a)　$\mathrm{cov}(\mathbf{L}_i'\mathbf{U}_1,\mathbf{L}_j'\mathbf{U}_1)=1$ ($i=j$ のとき)，$=0$ ($i\neq j$ のとき)

　　(b)　$\mathrm{cov}(\mathbf{M}_i'\mathbf{U}_2,\mathbf{M}_j'\mathbf{U}_2)=1$ ($i=j$ のとき)，$=0$ ($i\neq j$ のとき)

すなわち，$\mathbf{U}_1$ から求められる正準変数はすべて無相関で，標準偏差は1であり，$\mathbf{U}_2$ から導かれる正準変数も同様である．

(a) と (b) の結果は，それぞれ，(8f. 1. 7) と (8f. 1. 5) の式 $\mathscr{L}'\Sigma_{11}\mathscr{L}=\mathbf{I}$，$\mathscr{M}'\Sigma_{22}\mathscr{M}=\mathbf{I}$ から導かれる．

(iii)　(a)　$\mathrm{cov}(\mathbf{L}_i'\mathbf{U}_1,\mathbf{M}_i'\mathbf{U}_2)=\rho_i\neq 0, \quad i=1,\cdots,k$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad = 0, \quad i>k$

　　(b)　$\mathrm{cov}(\mathbf{L}_i'\mathbf{U}_1,\mathbf{M}_j'\mathbf{U}_2)=0, \quad i\neq j$

ただし，$k=R(\Sigma_{12})$．

(8f. 1. 2) の式の1つ

$$\Sigma_{12}\mathbf{M}_i=\rho_i\Sigma_{11}\mathbf{L}_i$$

を考える．これに $\mathbf{L}_i'$ をかけると，

$$\mathbf{L}_i'\Sigma_{12}\mathbf{M}_i=\rho_i\mathbf{L}_i'\Sigma_{11}\mathbf{L}_i=\rho_i$$

を得る．また，$\mathbf{L}_j'$ をかけると，

$$\mathbf{L}_j'\Sigma_{12}\mathbf{M}_i=\rho_i\mathbf{L}_j'\Sigma_{11}\mathbf{L}_i=0$$

を得る．一方，$\mathbf{L}_i'\Sigma_{12}\mathbf{M}_i=\mathrm{cov}(\mathbf{L}_i'\mathbf{U}_1,\mathbf{M}_i'\mathbf{U}_2)$ であり，$\mathbf{L}_j'\Sigma_{12}\mathbf{M}_i=\mathrm{cov}(\mathbf{L}_j'\mathbf{U}_1,\mathbf{M}_i'\mathbf{U}_2)$ であるから，求める結果が証明された．

このようにして，$\mathbf{U}_1,\mathbf{U}_2$ の同時変換 $\mathscr{L}'\mathbf{U}_1,\mathscr{M}'\mathbf{U}_2$ により，変換後の変数 $(\mathbf{U}_1'\mathscr{L}\,\vdots\,\mathbf{U}_2'\mathscr{M})$ の分散共分散行列は

$$\begin{pmatrix}\mathbf{I}_r & \mathbf{R}\\ \mathbf{R}' & \mathbf{I}_s\end{pmatrix}$$

という簡単な形になることがわかる．ただし，$\mathbf{I}_r,\mathbf{I}_s$ は次数 r,s の単位行列，$\mathbf{R}$ は $r\times s$ 行列でその最初の k 個の対角要素は $\rho_1,\cdots,\rho_k$ で残りの要素はすべて零である．

(iv)　$\mathbf{q}_i$ を $\mathscr{L}'^{-1}$ の第 i 列，$\mathbf{m}_i$ を $\mathscr{M}'^{-1}$ の第 i 列とすると，次の式が得られる．

(a)　$\Sigma_{11}=(\mathscr{L}^{-1})'(\mathscr{L}^{-1})=\mathbf{q}_1\mathbf{q}_1'+\cdots+\mathbf{q}_r\mathbf{q}_r'$,

(b)　$\Sigma_{22}=(\mathscr{M}^{-1})'(\mathscr{M}^{-1})=\mathbf{m}_1\mathbf{m}_1'+\cdots+\mathbf{m}_s\mathbf{m}_s'$,

(c)　$\Sigma_{12}=(\mathscr{L}^{-1})'\mathbf{R}(\mathscr{M}^{-1})=\rho_1\mathbf{q}_1\mathbf{m}_1'+\cdots+\rho_k\mathbf{q}_k\mathbf{m}_k'$.

(a), (b) および (c) の結果は，変換後の変数 $\mathscr{L}'\mathbf{U}_1,\mathscr{M}'\mathbf{U}_2$ の分散共分散行列を考えると導かれる．

$$D\begin{pmatrix}\mathscr{L}'\mathbf{U}_1\\ \hdashline \mathscr{M}'\mathbf{U}_2\end{pmatrix}=\begin{pmatrix}\mathscr{L}'\\ \hdashline \mathscr{M}'\end{pmatrix}D\begin{pmatrix}\mathbf{U}_1\\ \hdashline \mathbf{U}_2\end{pmatrix}(\mathscr{L}\,\vdots\,\mathscr{M})=\begin{pmatrix}\mathscr{L}'\Sigma_{11}\mathscr{L} & \mathscr{L}'\Sigma_{12}\mathscr{M}\\ \mathscr{M}'\Sigma_{21}\mathscr{L} & \mathscr{M}'\Sigma_{22}\mathscr{M}\end{pmatrix}$$
$$=\begin{pmatrix}\mathbf{I}_r & \mathbf{R}\\ \mathbf{R}' & \mathbf{I}_s\end{pmatrix}.$$

8f. 3　共通因子の有効個数

正準相関の現実的な意味を理解するために，次のような場面を考えよう．観測された計測値は，いくつかの無相関な仮説的因子（観測不可能な計測値）に線形従属であるとする．そのうち若干の因子は，$\mathbf{U}_1$ と $\mathbf{U}_2$ の両方の成分に影響しているが，他の因子は $\mathbf{U}_1$ または $\mathbf{U}_2$ のどちらかの成分にだけ影響しているかも知れない．このとき，前者を $\mathbf{U}_1,\mathbf{U}_2$ の共通因子とよび，後者を特殊因子とよぶ．すると，変数 $\mathbf{U}_1,\mathbf{U}_2$ の構造は，

$$\mathbf{U}_1=\mathbf{A}_1\mathbf{F}+\mathbf{G}_1, \qquad \mathbf{U}_2=\mathbf{A}_2\mathbf{F}+\mathbf{G}_2 \qquad (8f.3.1)$$

という形であらわされる．ただし，$\mathbf{F}'=(F_1,\cdots,F_m)$ は共通因子ベクトル，$\mathbf{A}_1$ と $\mathbf{A}_2$ は結合係数の行列，$\mathbf{G}_1,\mathbf{G}_2$ は特殊因子の効果の和である．定義により，

$$\mathrm{cov}(\mathbf{F},\mathbf{G}_1)=0, \quad \mathrm{cov}(\mathbf{F},\mathbf{G}_2)=0, \quad \mathrm{cov}(\mathbf{G}_1,\mathbf{G}_2)=0 \qquad (8f.3.2)$$

である．一般性を失うことなく，$D(\mathbf{F})=\mathbf{I}$ とすると．(8f. 3. 2) を用いて，

$$D(\mathbf{U}_1)=\Sigma_{11}=\mathbf{A}_1\mathbf{A}_1'+D(\mathbf{G}_1), \qquad D(\mathbf{U}_2)=\Sigma_{22}=\mathbf{A}_2\mathbf{A}_2'+D(\mathbf{G}_2)$$
$$C(\mathbf{U}_1,\mathbf{U}_2)=\Sigma_{12}=C(\mathbf{A}_1\mathbf{F},\mathbf{A}_2\mathbf{F})=\mathbf{A}_1\mathbf{A}_2' \qquad (8f.3.3)$$

を得る．これから，$\mathbf{U}_1, \mathbf{U}_2$ の関連性は，共通因子の影響だけによることがわかる．

当然のことながら，(8f. 3. 2)，(8f. 3. 3) の制約の下で，(8f. 3. 1) という表現が可能であるような共通因子の最小数 m はいくつであるかという疑問が生じる．この数を，共通因子の有効個数とよぶ．

共通因子の有効個数は $R(\Sigma_{12})$ に等しく，それはまた，(8f. 1. 4) または (8f. 1. 6) の非零根の数と同じであることを示そう．

(8f. 3. 3) から，$\Sigma_{12} = \mathbf{A}_1\mathbf{A}_2' \Rightarrow R(\Sigma_{12}) \leqslant \min(R(\mathbf{A}_1), R(\mathbf{A}_2))$．一方，$m \geqslant \max(R(\mathbf{A}_1), R(\mathbf{A}_2))$．したがって，$m \geqslant R(\Sigma_{12})$．事実，$m$ はこの条件の下で任意に選ぶことができる．しかしながら，$m = R(\Sigma_{12})$ が許容できる値である．この最後の結果は，$k = \operatorname{rank}\Sigma_{12}$ であることに注意し，

$$\mathbf{A}_1 = (\sqrt{\rho_1}\mathbf{q}_1 \vdots \cdots \vdots \sqrt{\rho_k}\mathbf{q}_k), \qquad \mathbf{A}_2 = (\sqrt{\rho_1}\mathbf{m}_1 \vdots \cdots \vdots \sqrt{\rho_k}\mathbf{m}_k)$$

に選ぶことによって，式

$$\Sigma_{12} = \rho_1\mathbf{q}_1\mathbf{m}_1' + \cdots + \rho_k\mathbf{q}_k\mathbf{m}_k' \tag{8f. 3. 4}$$

が成立することから得られる．$\mathbf{A}_1, \mathbf{A}_2$ のこのような選択が妥当であることを証明するためには，

$$D(\mathbf{G}_1) = \Sigma_{11} - \mathbf{A}_1\mathbf{A}_1' \quad \text{および} \quad D(\mathbf{G}_2) = \Sigma_{22} - \mathbf{A}_2\mathbf{A}_2'$$

が非負定符号であることを示さねばならない [(8f. 3. 3) 参照]．しかし

$$\Sigma_{11} - \mathbf{A}_1\mathbf{A}_1' = \mathbf{q}_1\mathbf{q}_1' + \cdots + \mathbf{q}_r\mathbf{q}_r' - (\rho_1\mathbf{q}_1\mathbf{q}_1' + \cdots + \rho_k\mathbf{q}_k\mathbf{q}_k')$$

で，$\rho_i \leqslant 1$，$i = 1, \cdots, k$ であるから，$\Sigma_{11} - \mathbf{A}_1\mathbf{A}_1'$ は非負定符号であることがわかる．$\Sigma_{22} - \mathbf{A}_2\mathbf{A}_2'$ についても同じことがいえる．

分散共分散行列の推定値から共通因子の有効個数を推定するには重大な問題が生じる．Σ_{12} の推定値である $\mathbf{S}_{12}$ の階数は，$\operatorname{rank}\Sigma_{12} < \min(r, s)$ のときでも，確率 1 で $\min(r, s)$ になるであろう．したがって，$\mathbf{S}_{12}$ の階数から，共通因子の有効個数について何も推測することはできない．ただ 1 つの方法は，式 (8f. 1. 4) の $\Sigma_{11}, \Sigma_{12}, \Sigma_{22}$ のかわりに，推定値 $\mathbf{S}_{11}, \mathbf{S}_{12}, \mathbf{S}_{22}$ を用いた固有方程式

$$|\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} - r^2\mathbf{S}_{22}| = 0 \tag{8f. 3. 5}$$

の根をすべて求め，それらの根の大きさを吟味することである．(8f. 1. 4) の根 $\rho_1^2, \rho_2^2, \cdots$ のうち k 個だけが非零のときは，仮説的な根が零に対応する (8f. 3. 5) の根 $r_{k+1}^2, r_{k+2}^2, \cdots$ は小さくなりやすいであろう．したがって，小さい方の根の有意性を検定することによって，共通因子の有効個数について推測することが可能になるであろう．

しかしながら，現実のどのような問題でも，共通因子の有効個数は非常に大きくなりがちであるから，われわれの提起した問題も，示唆した推定の方法も意味がなくなる．ではあるが，その場合でも正しいといえるのは，2 組の変数のあいだの関連性の大部分がより

少数の共通因子で説明できるということである．そのような因子を優越共通因子とよび，その数を推定する問題は意味のある方法によって，また必要なら数学的に厳密に定式化することができる．

式 (8f. 3. 4)

$$\Sigma_{12} = \rho_1\mathbf{q}_1\mathbf{m}_1' + \cdots + \rho_k\mathbf{q}_k\mathbf{m}_k'$$

は，Σ_{12} への寄与の大部分は，比較的大きい ρ_i をもつ項によって占められることを示すから，優越因子の数は ρ_i のなかの大きな根の数によって定まる．このとき，仮説的優越根の数，つまり，優越因子の数についての推測を行うのに，(式 (8f. 3. 5) の) 根の推定値を用いてよいであろう．

8f. 4　因 子 分 析

$\mathbf{U}$ は次の構造をもつ p 次元 r. v. とする，

$$\mathbf{U} = \mathbf{AF} + \mathbf{G} \tag{8f. 4. 1}$$

ただし，$\mathbf{F}' = (F_1, \cdots, F_m)$ と $\mathbf{G}' = (G_1, \cdots, G_p)$ はすべて互いに無相関な変数で，$\mathbf{A}$ は (未知) 定数の $p \times m$ 行列である．**8f. 3** の術語を用いて，$F_1, \cdots, F_m$ を共通因子，$G_1, \cdots, G_p$ を特殊因子とよぶ．一般性を失うことなく，$V(G_i) = \delta_i$，$V(F_i) = 1$ とおくことができる．このとき，

$$D(\mathbf{U}) = \Sigma = \mathbf{AA}' + \Delta \tag{8f. 4. 2}$$

と書ける．ただし，Δ はその対角要素が $\delta_1, \cdots, \delta_p$ の対角行列である．2 組の変数のあいだの関係を問題にしたときと同様に，(8f. 4. 2) の制約の下で，式 (8f. 4. 1) が成り立つような m の最小値を，共通因子の有効個数と定義しよう．ここで述べた問題は，所与の状態ではいつも共通因子の有効個数が大きいのがふつうであるから，あまり現実的ではない．しかし，1 組の変数のあいだの相互の相関を説明するのに，いくつの共通因子が重要であるかを調べることは妥当であろう．そのような問題への適切な接近方法は著者 (Rao (1955b)) によって示された．この問題に関しては尨大な文献があり，興味のある読者は，Lawley and Maxwell (1963) の最近の本を参照するとよい．いくつかの母集団における同時因子分析の問題は Rasch (1953) が研究しており，また，Rao (1966a, 1967d, 1969) も参照されたい．

8g　確率変数の正規直交基底

8g. 1　グラム＝シュミット基底

$U_1, \cdots, U_p$ は p 個の (1 次元) r. v. で，$E(U_i) = 0$，分散共分散行列 Σ の階数は $m \leqslant p$

とする．$\mathcal{M}(U_1,\cdots,U_p)$ あるいは簡単に $\mathcal{M}(\mathbf{U})$ と書いて，実係数 c_i をもつすべての r. v. $c_1U_1+\cdots+c_pU_p$ の集合をあらわす．定義によって $\mathcal{M}(\mathbf{U})$ はベクトル空間である．$\mathcal{M}(\mathbf{U})$ の任意の2つの要素 Y_1, Y_2 の内積を $\mathrm{cov}(Y_1,Y_2)$ と定義する．このとき，Y_1 のノルムは標準偏差となる，つまり $\|Y_1\|=[V(Y_1)]^{1/2}$．このとき，1次元 r. v. のベクトル空間 $\mathcal{M}(\mathbf{U})$ はノルム空間となる．

ベクトル空間の理論から，$\mathcal{M}(\mathbf{U})$ は，単位ノルムをもち，互いの内積は零の o. n. b.（正規直交基底）$G_1,\cdots,G_m$ をもつ（つまり，G_i は標準偏差1で，共分散は零の r. v. である）ことがわかる．このとき，$\mathcal{M}(\mathbf{U})$ のどの要素（r. v.）も，要素（r. v. である）$G_1,\cdots,G_m$ の1次結合としてあらわされる．特に

$$U_i = a_{i1}G_1+\cdots+a_{im}G_m, \qquad i=1,\cdots,p \tag{8g.1.1}$$

となる．$\mathbf{G}'=(G_1,\cdots,G_m)$ と書くと，(8g.1.1) の関係は，行列記号で，

$$\mathbf{U}=\mathbf{AG}, \qquad \mathbf{\Sigma}=\mathbf{AA}' \tag{8g.1.2}$$

とあらわされる．反対に，$\mathcal{M}(\mathbf{U})=\mathcal{M}(\mathbf{G})$ であるから，$\mathbf{G}$ は $\mathbf{U}$ を用いて，

$$\mathbf{G}=\mathbf{BU}, \qquad \mathbf{I}=\mathbf{B\Sigma B}' \tag{8g.1.3}$$

とあらわすことができる．

さらに，ベクトル空間の理論から，$\mathcal{M}(\mathbf{U})$ の次元は要素 $U_1,\cdots,U_p$ の内積の行列，つまりこの場合は Σ の階数に一致することが知られる．したがって，o. n. b. における変数の数は $m=R(\Sigma)$ である．o. n. b. は一意ではない．しかしながら，統計的に興味のあるのは，そのいくつかの特殊の場合で，それについて以下で考えよう．

射影としての線形予測子　U_i の $\mathcal{M}(U_1,\cdots,U_{i-1})$ への射影を $P(U_i)$ とする．定義によれば，それは，$U_1,\cdots,U_{i-1}$ の1次関数で，

$$\left\|U_i-\sum_1^{i-1}b_rU_r\right\|^2 = E\left(U_i-\sum b_rU_r\right)^2 \tag{8g.1.4}$$

を最小にするものである．したがって，$P(U_i)$ は，$U_1,\cdots,U_{i-1}$ に基づく U_i の最小平均2乗誤差線形予測子である．さらに，$U_i-P(U_i)$ は

$$\mathcal{M}(U_1,\cdots,U_{i-1})$$

に直交するから，最良1次関数の係数は，$(U_i-\sum b_rU_r)$ と U_s $(s=1,\cdots,i-1)$ との内積がすべて零となるという条件をあらわすことによって求められる．内積は共分散であるから，その式は

$$\mathrm{cov}\left(U_s,\ U_i-\sum_1^{i-1}b_rU_r\right)=0, \qquad s=1,\cdots,i-1 \tag{8g.1.5}$$

となり，これから，

$$\begin{gathered}\mathrm{cov}[U_s,\ U_i-P(U_i)]=0, \qquad s=1,\cdots,i-1\\ \mathrm{cov}[U_i-P(U_i),\ U_j-P(U_j)]=0, \qquad i\neq j\end{gathered} \tag{8g.1.6}$$

となることがわかる．もし，残差 $U_i-P(U_i)$ を $U_{i\cdot12\cdots(i-1)}$ であらわすと，(8g.1.6) は，残差

$$U_1,\quad U_{2\cdot1},\quad U_{3\cdot12},\quad\cdots,\quad U_{p\cdot12\cdots(p-1)} \tag{8g.1.7}$$

がすべて互いに無相関であるということと等価になる．

グラム-シュミット o. n. b.　残差 $U_{i\cdot12\cdots(i-1)}$ の標準偏差を $t_{ii}=\|U_i-P(U_i)\|$ とし，

$$\begin{aligned} t_{11}G_1 &= U_1 &&= U_1\\ t_{22}G_2 &= U_2-P(U_1) &&= U_2-t_{21}G_1\\ \cdots &\quad \cdots &&\quad \cdots \\ t_{pp}G_p &= U_p-P(U_p) &&= U_p-t_{p1}G_1-\cdots-t_{p,p-1}G_{p-1}\end{aligned} \tag{8g.1.8}$$

を考える．U_i の $\mathcal{M}(U_1,\cdots,U_{i-1})$ 上への射影は $G_1,\cdots,G_{i-1}$ によってあらわされ，その係数はグラム-シュミットの直交化手順［(ii), 1a. 4］のときと同様に，次々に求めて行くことができる点に注意せよ．$t_{ii}=0$ なら，U_i はそれ以前の変数によって完全に定まる．そのときは，$G_i=0$ とおく．それ以外の場合は，$\|G_i\|=1$ となるように

$$G_i=[U_i-P(U_i)]\div t_{ii}=U_{i\cdot12\cdots(i-1)}\div t_{ii}$$

とおけば，(8g.1.7) を用いて，非零の t_{ii} に対応する G_i は o. n. b. をつくる．(8g.1.8) の逆の関係は，

$$\begin{aligned} U_1 &= t_{11}G_1\\ U_2 &= t_{21}G_1+t_{22}G_{22}\\ \cdots &\quad \cdots \quad \cdots\\ U_p &= t_{p1}G_1+t_{p2}G_2+\cdots+t_{pp}G_p\end{aligned} \tag{8g.1.9}$$

となり，$\mathbf{T}$ を(8g.1.9) に示されるような下三角行列とすると，$\mathbf{U}=\mathbf{TG}$ と書ける．このとき，

$$\mathbf{\Sigma}=E(\mathbf{UU}')=\mathbf{T}E(\mathbf{GG}')\mathbf{T}'=\mathbf{TT}' \tag{8g.1.10}$$

となる．式 (8g.1.10) は，$\mathbf{T}$ が，簡単でよく知られた平方根標準化法を Σ に施すことによって得られる行列そのものであることを示す．$R(\Sigma)=p$ のときは，$\mathbf{T}$ の対角要素はすべて非零である．それ以外のときには，$p-m$ 個の対角要素とそれに対応する列が零になり，(8g.1.9) の関係は G_i 変数を m 個だけ含むことになる．

実際には，（相関のある変数）$\mathbf{U}$ を（互いに無相関な変数）$\mathbf{G}$ によってあらわすことと，$\mathbf{G}$ を $\mathbf{U}$ によってあらわすという2つの関係は，必要なら，$\mathbf{\Sigma}$ の右に単位行列 $\mathbf{I}$ をつけた

$$\mathbf{\Sigma}\ \vdots\ \mathbf{I} \tag{8g.1.11}$$

に平方根法を適用することによって，同時に得ることができる．平方根法による標準化を行うと，

$$\mathbf{T}'\ \vdots\ \mathbf{W} \tag{8g.1.12}$$

が得られる，ここで，$\mathbf{T}'$ は上三角行列，$\mathbf{W}$ は下三角行列となる．このとき，$\mathbf{G}$ から $\mathbf{U}$ へ

の変換と，$\mathbf{U}$ から $\mathbf{G}$ への変換は，

$$\mathbf{U}=\mathbf{TG},\qquad \mathbf{G}=\mathbf{WU} \tag{8g.1.13}$$

によって与えられる．

8g.2 主成分分析

分散共分散行列 $\boldsymbol{\Sigma}$ をもつ p 次元 r.v. である $\mathbf{U}$ を考える．$\lambda_1\geqslant\cdots\geqslant\lambda_p$ を $\boldsymbol{\Sigma}$ の固有値 $\mathbf{P}_1,\cdots,\mathbf{P}_p$ を対応する固有ベクトルとする．このとき，(1c.3.5) で，

$$\boldsymbol{\Sigma}=\lambda_1\mathbf{P}_1\mathbf{P}_1'+\cdots+\lambda_p\mathbf{P}_p\mathbf{P}_p',\qquad \mathbf{I}=\mathbf{P}_1\mathbf{P}_1'+\cdots+\mathbf{P}_p\mathbf{P}_p' \tag{8g.2.1}$$

$$\mathbf{P}_i'\boldsymbol{\Sigma}\mathbf{P}_i=\lambda_i,\qquad \mathbf{P}_i'\boldsymbol{\Sigma}\mathbf{P}_j=0,\qquad i\neq j \tag{8g.2.2}$$

が示された．$\mathbf{U}$ を変換した r.v. である，

$$Y_i=\mathbf{P}_i'\mathbf{U},\qquad i=1,\cdots,p$$

を考える．$\mathbf{Y}$ を新しい r.v. のベクトル，$\mathbf{P}$ を列ベクトルが $\mathbf{P}_1,\cdots,\mathbf{P}_p$ の直交行列とすると，$\mathbf{Y}$ は $\mathbf{U}$ から $\mathbf{Y}=\mathbf{P}'\mathbf{U}$ という直交変換によって求められる．r.v. Y_i を $\mathbf{U}$ の第 i 主成分と呼ぶ．以下では，主成分のもつ若干の性質とその解釈について吟味しよう．

(i) 主成分はすべて互いに無相関である．第 i 主成分の分散は λ_i である．

これらの結果は (8g.2.2) からわかる．

$$V(\mathbf{P}_i'\mathbf{U})=\mathbf{P}_i'\boldsymbol{\Sigma}\mathbf{P}_i=\lambda_i$$

$$\mathrm{cov}(\mathbf{P}_i'\mathbf{U},\mathbf{P}_j'\mathbf{U})=\mathbf{P}_i'\boldsymbol{\Sigma}\mathbf{P}_j=0,\qquad i\neq j$$

それゆえ，$\mathbf{Y}=\mathbf{P}'\mathbf{U}$ という1次変換は，互いに相関のある1組の変数を，直交変換によって，互いに無相関な1組に帰着させるものである．

(ii) $\lambda_i\neq0$ にたいして，$G_i=\lambda_i^{-1/2}\mathbf{P}_i'\mathbf{U}$ とし，さらに r を $\boldsymbol{\Sigma}$ の階数とすると，$\boldsymbol{\Sigma}$ のはじめの r 個の固有値だけが非零となる．このとき，$G_1,\cdots,G_r$ は r.v. $\mathbf{U}$ の o.n.b. となる．

(iii) $\mathbf{B}$ を $\|\mathbf{B}\|=1$ であるような任意のベクトルとする．このとき，$V(\mathbf{B}'\mathbf{U})$ は $\mathbf{B}=\mathbf{P}_1$ のとき最大となり，その最大分散は λ_1 となる．

$V(\mathbf{B}'\mathbf{U})=\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B}$ であるから，$\|\mathbf{B}\|=1$ の条件下で，$\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B}$ の最大値を求めればよい．一方，その最大値は $\mathbf{B}=\mathbf{P}_1$ のときに得られることを (1f.2.1) に示した．

実際，**1f.2** で証明した諸命題の結果として次を得る．

(iv) (a) $\displaystyle\min_{\|\mathbf{B}\|=1}V(\mathbf{B}'\mathbf{U})=\lambda_p=V(\mathbf{P}_p'\mathbf{U})$

(b) $\displaystyle\max_{\|\mathbf{B}\|=1,\mathbf{B}\perp\mathbf{P}_1,\cdots,\mathbf{P}_{i-1}}V(\mathbf{B}'\mathbf{U})=\lambda_i=V(\mathbf{P}_i'\mathbf{U})$

(c) $\displaystyle\min_{S_{i-1}}\max_{\|\mathbf{B}\|=1,\mathbf{B}\perp S_{i-1}}V(\mathbf{B}'\mathbf{U})=\lambda_i=V(\mathbf{P}_i'\mathbf{U})$

ただし，S_{i-1} は E_p における $i-1$ 次元部分空間とする．

(d) $\mathbf{B}_1,\cdots,\mathbf{B}_k$ を E_p における1組の正規直交ベクトルとすると，

$$\begin{aligned}\lambda_1+\cdots+\lambda_k&=\max_{\mathbf{B}_1,\cdots,\mathbf{B}_k}[V(\mathbf{B}_1'\mathbf{U})+\cdots+V(\mathbf{B}_k'\mathbf{U})]\\&=V(\mathbf{P}_1'\mathbf{U})+\cdots+V(\mathbf{P}_k'\mathbf{U}).\end{aligned}$$

(v) $\mathbf{B}_1'\mathbf{U},\cdots,\mathbf{B}_k'\mathbf{U}$ を $\mathbf{U}$ の k 個の1次関数，σ_i^2 を $\mathbf{B}_1'\mathbf{U},\cdots,\mathbf{B}_k'\mathbf{U}$ に基づく最良線形予測子によって U_i を予測するときの残差分散とする．このとき

$$\min_{\mathbf{B}_1,\cdots,\mathbf{B}_k}\sum_{i=1}^{p}\sigma_i^2$$

は，$\mathbf{B}_1'\mathbf{U},\cdots,\mathbf{B}_k'\mathbf{U}$ の組が $\mathbf{P}_1'\mathbf{U},\cdots,\mathbf{P}_k'\mathbf{U}$ の組に同値であるとき，すなわち，各 $\mathbf{B}_i'\mathbf{U}$ が最初の k 個の主成分の1次結合であらわされるときに達成される．

一般性を失うことなく，$\mathbf{B}_1'\mathbf{U},\cdots,\mathbf{B}_k'\mathbf{U}$ を，これと等価な，互いに無相関でそれぞれ分散1をもつ関数の組におきかえることができる．また関数 $\mathbf{B}_1'\mathbf{U},\cdots,\mathbf{B}_k'\mathbf{U}$ は最適解になるとき1次独立であるべきことも明らかである．$\mathbf{B}_1'\mathbf{U},\cdots,\mathbf{B}_k'\mathbf{U}$ に基づいて U_i を予測するときの残差分散は，

$$\begin{aligned}\sigma_i^2&=\sigma_{ii}-[\mathrm{cov}(U_i,\mathbf{B}_1'\mathbf{U})]^2-\cdots-[\mathrm{cov}(U_i,\mathbf{B}_k'\mathbf{U})]^2\\&=\sigma_{ii}-(\mathbf{B}_1'\boldsymbol{\Sigma}_i)^2-\cdots-(\mathbf{B}_k'\boldsymbol{\Sigma}_i)^2\\&=\sigma_{ii}-\mathbf{B}_1'\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_i'\mathbf{B}_1-\cdots-\mathbf{B}_k'\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_i'\mathbf{B}_k\end{aligned}$$

となる．ただし，$\boldsymbol{\Sigma}_i$ は $\boldsymbol{\Sigma}$ の第 i 列，σ_{ii} は U_i の分散である．すると，

$$\begin{aligned}\sum_{i=1}^{p}\sigma_i^2&=\sum_{i=1}^{p}\sigma_{ii}-\mathbf{B}_1'\Big(\sum_{i=1}^{p}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_i'\Big)\mathbf{B}_1-\cdots-\mathbf{B}_k'\Big(\sum_{i=1}^{p}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_i'\Big)\mathbf{B}_k\\&=\mathrm{trace}\,\boldsymbol{\Sigma}-\mathbf{B}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}\mathbf{B}_1-\cdots-\mathbf{B}_k'\boldsymbol{\Sigma}\boldsymbol{\Sigma}\mathbf{B}_k\end{aligned}$$

となる．$\sum\sigma_i^2$ を最小にするためには

$$\mathbf{B}_1'\boldsymbol{\Sigma}\boldsymbol{\Sigma}\mathbf{B}_1+\cdots+\mathbf{B}_k'\boldsymbol{\Sigma}\boldsymbol{\Sigma}\mathbf{B}_k \tag{8g.2.3}$$

を最大にすればよい，ただし，

$$\mathbf{B}_i'\boldsymbol{\Sigma}\mathbf{B}_i=1,\qquad \mathbf{B}_i'\boldsymbol{\Sigma}\mathbf{B}_j=0,\qquad i\neq j \tag{8g.2.4}$$

という条件（すなわち，$\mathbf{B}_1'\mathbf{U},\cdots,\mathbf{B}_k'\mathbf{U}$ は互いに無相関で分散は1）がつく．そのような場合の $\mathbf{B}_i$ の最適な選択は，第1章末の問題 **22** で示したように，

$$\mathscr{M}(\mathbf{B}_1,\cdots,\mathbf{B}_k)=\mathscr{M}(\mathbf{P}_1,\cdots,\mathbf{P}_k) \tag{8g.2.5}$$

を成立させるものである．ここに $\mathbf{P}_1,\cdots,\mathbf{P}_k$ は

$$|\boldsymbol{\Sigma}-\lambda\mathbf{I}|=0 \tag{8g.2.6}$$

の最初の k 個の固有ベクトルである．

(v) の結果は，主成分について興味深い解釈を与える．いま，p 次元 r.v. である $\mathbf{U}$ を情報の大きな損失なしに，$k<p$ 個の1次関数におきかえたいとしよう．最良の k 個の1次関数はどのようにして選ぶべきであるか？ k 個の1次関数をどのように選んでも，その効率は，その k 個の1次関数によって，もとの p 個の変数がどの程度再現できるかに依存する．変数 U_i を再現させる1つの方法は，k 個の1次関数に基づくその最良線形予測

子を定めるやり方であり，その場合の予測の効率は残差分散 σ_i^2 によって測られる．予測効率の全体としての尺度は $\sum\sigma_i^2$ となる．$\sum\sigma_i^2$ が最小となるような最適な1次関数の選択は，$\mathbf{U}$ の最初の k 個の主成分を選ぶことである．全体としての予測効率についてのずっと強い尺度をとっても同じ結果が得られることを (vi) で示そう．

k 個の1次関数 $\mathbf{B}_1'\mathbf{U}, \cdots, \mathbf{B}_k'\mathbf{U}$ を，変換 $\mathbf{Y} = \mathbf{B}'\mathbf{U}$ であらわすことにしよう．ただし，$\mathbf{B}$ は $p \times k$ 行列で，$\mathbf{B}_1, \cdots, \mathbf{B}_k$ はその列である．すると，$\mathbf{U}$ と $\mathbf{Y}$ との同時分散共分散行列は

$$\begin{pmatrix} \Sigma & \Sigma\mathbf{B} \\ \mathbf{B}'\Sigma & \mathbf{B}'\Sigma\mathbf{B} \end{pmatrix} \tag{8g. 2. 7}$$

となり，$\mathbf{U}$ からその $\mathbf{Y}$ による最良線形予測子を差引いた残差の分散共分散行列は，

$$\Sigma - \Sigma\mathbf{B}(\mathbf{B}'\Sigma\mathbf{B})^{-1}\mathbf{B}'\Sigma \tag{8g. 2. 8}$$

となる．(8g. 2. 8) の各要素の値が小さければ小さいほど，$\mathbf{Y}$ による予測効率は大きくなる．(v) では

$$\mathrm{trace}(\Sigma - \Sigma\mathbf{B}(\mathbf{B}'\Sigma\mathbf{B})^{-1}\mathbf{B}'\Sigma) \tag{8g. 2. 9}$$

を最小にするという方法で $\mathbf{B}$ を求めた．より強い結果を次に証明しよう．

(vi)　ユークリッド・ノルム

$$\|\Sigma - \Sigma\mathbf{B}(\mathbf{B}'\Sigma\mathbf{B})^{-1}\mathbf{B}'\Sigma\| \tag{8g. 2. 10}$$

は，ベクトル $\mathbf{B}_1, \cdots, \mathbf{B}_k$ の組が固有ベクトル $\mathbf{P}_1, \cdots, \mathbf{P}_k$ の組と等価のとき，最小になる．

このようにして，全予測効率の2つの測度，つまり，

(a)　$\mathrm{trace}(\Sigma - \Sigma\mathbf{B}(\mathbf{B}'\Sigma\mathbf{B})^{-1}\mathbf{B}'\Sigma)$

(b)　$\|\Sigma - \Sigma\mathbf{B}(\mathbf{B}'\Sigma\mathbf{B})^{-1}\mathbf{B}'\Sigma\|$

は，ベクトル $\mathbf{B}_i$ の最適選択に関して同じ結果をもたらす．

(vi) の結果を証明するためには，(1f. 2. 10) を適用すればよい．そこでは，(8g. 2. 10) の最小値が

$$\Sigma\mathbf{B}(\mathbf{B}'\Sigma\mathbf{B})^{-1}\mathbf{B}'\Sigma = \lambda_1\mathbf{P}_1\mathbf{P}_1' + \cdots + \lambda_k\mathbf{P}_k\mathbf{P}_k' \tag{8g. 2. 11}$$

のとき得られることを示した．(8g. 2. 11) の等式を満足させるには，$\mathbf{B}_i = \mathbf{P}_i$ と選べばよいことが容易に証明できる．

主成分の推定およびそれに関連する有意性検定の研究には，Kshirsagar (1961, 1972) を参照すればよい．Kshirsagar (1972) の本は，多変量分布問題についての良い解説を含んでいる．

補足と問題

1　$\mathbf{U}_1, \cdots, \mathbf{U}_n$ を p 次元ユークリッド空間 E_p における n 個の点とする．点 $\bar{\mathbf{U}} = n^{-1}\sum\mathbf{U}_i$ はこれらの点の重心である．$\mathscr{U}$ と書いて，i 番目の列が $\mathbf{U}_i - \bar{\mathbf{U}}$ である $p \times n$

行列をあらわす．行列 $\mathscr{U}\mathscr{U}'$ と $\mathscr{U}'\mathscr{U}$ を考える．この2つの行列は同じ非零固有値をもつことが知られている．$\mathbf{P}_i$ と $\mathbf{Q}_i$ を固有値 $\lambda_i \neq 0$ に対応する $\mathscr{U}\mathscr{U}'$ と $\mathscr{U}'\mathscr{U}$ の固有ベクトルとすれば，$\sqrt{\lambda_i}\mathbf{Q}_i = \mathscr{U}'\mathbf{P}_i$，$\sqrt{\lambda_i}\mathbf{P}_i = \mathscr{U}\mathbf{Q}_i$ である．

1.1　E_p における点 $\mathbf{V}_1, \cdots, \mathbf{V}_n$ を，k 次元空間に存在し，かつ

$$\sum_{i=1}^{n}(\mathbf{U}_i - \mathbf{V}_i)'(\mathbf{U}_i - \mathbf{V}_i)$$

が最小になるように定めよ．［最適点 $\mathbf{V}_1^*, \cdots, \mathbf{V}_n^*$ が存在する k 次元空間 S_k は，点 $\bar{\mathbf{U}}$ と $\mathscr{U}\mathscr{U}'$ の最初の k 個の固有ベクトルである k 個の正規直交ベクトル $\mathbf{P}_1, \cdots, \mathbf{P}_k$ によって定められることを示せ．それより

$$\mathbf{V}_i^* = \bar{\mathbf{U}} + \mathbf{P}_1'(\mathbf{U}_i - \bar{\mathbf{U}})\mathbf{P}_1 + \cdots + \mathbf{P}_k'(\mathbf{U}_i - \bar{\mathbf{U}})\mathbf{P}_k \qquad i = 1, \cdots, n$$

を示せ．］さらに，$\mathscr{U}\mathscr{U}'$ の固有値を $\lambda_1, \cdots, \lambda_p$ とすれば，

$$\min\sum(\mathbf{U}_i - \mathbf{V}_i)'(\mathbf{U}_i - \mathbf{V}_i) = \lambda_{k+1} + \cdots + \lambda_p$$

であることを示せ．

1.2　**1.1** で定義された k 次元空間 S_k は，所与の点 $\mathbf{U}_1, \cdots, \mathbf{U}_n$ から S_k への垂線の2乗和が他のいかなる k 次元空間と比べても最小であるという意味で，これらの点に最も適合する k 次元空間であることを示せ．

1.3　$\mathbf{B}$ は $k \times p$ 行列で，$\mathbf{Y} = \mathbf{B}\mathscr{U}$ は E_p の点を E_k $(k < p)$ の点へ移す変換をあらわすとする．さらに，$\mathbf{Y}_i = \mathbf{B}\mathbf{U}_i$，$i = 1, \cdots, n$ は E_p の点 $\mathbf{U}_1, \cdots, \mathbf{U}_n$ を変換した点とする．〔$\mathbf{B}\mathbf{B}' = \mathbf{I}_k$ のもとでは，〕$\mathbf{B}'$ の列が $\mathscr{U}\mathscr{U}'$ の最初の k 個の固有ベクトル $\mathbf{P}_1, \cdots, \mathbf{P}_k$ であるときに，

$$\sum_{i,j=1}^{n}\sum(\mathbf{Y}_i - \mathbf{Y}_j)'(\mathbf{Y}_i - \mathbf{Y}_j)$$

が最大になることを示せ．上の和の最大値は，$\mathscr{U}\mathscr{U}'$ の固有値を λ_i とすれば，$2n(\lambda_1 + \cdots + \lambda_k)$ となることを示せ．

1.4　行列 $\mathscr{U}'\mathscr{U}$ を考える．この行列の第 i 対角要素は $\|\mathbf{U}_i - \bar{\mathbf{U}}\|^2$ で，これは点 $\mathbf{U}_i$ の重心からの距離の2乗である．その (i, j) 要素は，ベクトル $\mathbf{U}_i - \bar{\mathbf{U}}$ と $\mathbf{U}_j - \bar{\mathbf{U}}$ のつくる角の余弦とその距離 $\|\mathbf{U}_i - \bar{\mathbf{U}}\|$，$\|\mathbf{U}_j - \bar{\mathbf{U}}\|$ との積である．それゆえ，$\mathscr{U}'\mathscr{U}$ は点 $\mathbf{U}_1, \cdots, \mathbf{U}_n$ と重心とを結ぶ直線の距離と角度によって測られた，点の配置を特徴づける行列である．$\mathbf{Y} = \mathbf{B}\mathscr{U}$ は，問題 **1.3** におけるように，E_p の点の E_k への変換とする．変換された点の配置は，行列 $\mathbf{Y}'\mathbf{Y} = \mathscr{U}'\mathbf{B}'\mathbf{B}\mathscr{U}$ によってあらわされる．ユークリッド・ノルム $\|\mathscr{U}'\mathscr{U} - \mathscr{U}'\mathbf{B}'\mathbf{B}\mathscr{U}\|$ を尺度として，変換による配置のヒズミが最小になるように行列 $\mathbf{B}$ を決定せよ．

［ヒント：(1f. 2. 10) の結果を用いよ．それは $\mathbf{B}$ の最適な選択は，

$$\mathscr{U}'\mathbf{B}'\mathbf{B}\mathscr{U} = \lambda_1\mathbf{Q}_1\mathbf{Q}_1' + \cdots + \lambda_k\mathbf{Q}_k\mathbf{Q}_k'$$

をみたすものであることを示している．ただし，$\mathbf{Q}_1, \cdots, \mathbf{Q}_k$ は $\mathscr{U}'\mathscr{U}$ の固有ベクトルである．一方，$\mathbf{Q}_i = \lambda_i^{-1/2}\mathscr{U}'\mathbf{P}_i$ である．したがって

$$\mathscr{U}'\mathbf{B}'\mathbf{B}\mathscr{U} = \mathscr{U}'(\mathbf{P}_1\mathbf{P}_1' + \cdots + \mathbf{P}_k\mathbf{P}_k')\mathscr{U}$$

となり，よって，$\mathbf{B}'\mathbf{B} = \mathbf{P}_1\mathbf{P}_1' + \cdots + \mathbf{P}_k\mathbf{P}_k'$，すなわち $\mathbf{B}'$ の列は固有ベクトル $\mathbf{P}_1, \cdots, \mathbf{P}_k$ であることを教える．］

2　$u_1, \cdots, u_k$ と $v_1, \cdots, v_m$ をそれぞれ k 個と m 個の1次元確率変数の2組とする．そ

れらの同時分散共分散行列は，分割行列の形で，

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

と書ける．u_i の一部分または全体が変数 v_i の部分集合であってもよい．

2.1　1次関数 $b_1v_1+\cdots+b_mv_m$ を考える．u_i の $b_1v_1+\cdots+b_mv_m$ にたいする1次回帰による予測の残差分散は

$$\sigma_i^2=V(u_i)-\frac{[\mathrm{cov}(u_i, b_1v_1+\cdots+b_mv_m)]^2}{V(b_1v_1+\cdots+b_mv_m)}$$

である．$\sum_1^k \sigma_i^2$ を最小にするベクトル $(b_1, \cdots, b_m)$ は

$$|\Sigma_{21}\Sigma_{12}-\lambda\Sigma_{22}|=0$$

の最大根に対応する固有ベクトルであることを示せ．

2.2　$u_1, \cdots, u_k$ を予測するのに残差分散の和を最小にするという意味で，最良な $v_1, \cdots, v_m$ の r 個の1次関数は，**2.1** の固有方程式から得られる最初から r 個の固有ベクトルに対応するものであることを示せ．

2.3　$u_1, \cdots, u_k$ が $v_1, \cdots, v_m$ の部分集合であるときには，最初の k 個の固有ベクトルは $u_1, \cdots, u_k$ のみからなる1次関数をつくり，それらは $u_1, \cdots, u_k$ の主成分に相当することを示せ．

3　$u_1, \cdots, u_k$ を k 個の1次元確率変数とする．$w_1^2, \cdots, w_k^2$ が所与の重みであるとき，

$$\sum_1^k w_i^2\left\{V(u_i)-\frac{[\mathrm{cov}(u_i, b_1u_1+\cdots+b_ku_k)]^2}{V(b_1u_1+\cdots+b_ku_k)}\right\}$$

が最小になるように，1次関数 $b_1u_1+\cdots+b_ku_k$ を決定せよ．

4　$\mathbf{U}_1, \cdots, \mathbf{U}_n$ を E_p の n 個の点とする．n 個の点 $\mathbf{V}_1, \cdots, \mathbf{V}_n$ が k 次元空間に存在し，かつ

$$\sum_1^n(\mathbf{U}_i-\mathbf{V}_i)'\Lambda(\mathbf{U}_i-\mathbf{V}_i)$$

が最小となるようにこれらの点を決定せよ．ただし Λ は所与の正定符号行列とする．

5　p 変量正規母集団の平均ベクトルが指定された値をもつという仮説を，大きさ n の標本に基づいて検定するためのホテリングの T^2 検定は，尤度比原理によって得られることを示せ．

6　平均が $\mathfrak{d}_1, \mathfrak{d}_2$ で，共通の分散共分散行列 Σ をもつ，p 次元確率変数 $\mathbf{U}$ の2つの分布を考える．母集団間の Mahalanobis (1936) の距離 D^2 は

$$D^2=(\mathfrak{d}_1-\mathfrak{d}_2)'\Sigma^{-1}(\mathfrak{d}_1-\mathfrak{d}_2)$$

によって定義される．この D^2 は変数 $\mathbf{U}$ の1次変換にたいして不変であることを示せ．D^2 が平方和になるような1次変換を見出せ．

7　$\mathbf{U}_i \sim N_p(a_i\mathfrak{y}, \Sigma)$，$i=1, \cdots, n$ は互いに独立であるとする．

$$\mathbf{Q}=\sum \mathbf{U}_r\mathbf{U}_r'-\mathbf{Z}\mathbf{Z}' \quad と \quad \mathbf{Z}\mathbf{Z}'$$

は互いに独立に分布することを示せ．ただし，$\mathbf{Z}=\sum b_i\mathbf{U}_i$，$b_i=a_i/\sqrt{\sum a_i^2}$．

［ヒント：(8b.1.1) の記号より，$\mathbf{Z}=\mathscr{U}'\mathbf{B}$，$\mathbf{B}'=(b_1, \cdots, b_n)$，$\mathbf{Q}=\mathscr{U}'\mathscr{U}-\mathscr{U}'\mathbf{B}\mathbf{B}'\mathscr{U}=\mathscr{U}'\mathbf{A}\mathscr{U}$ である．$\mathbf{AB}=0$ を証明し，［(ⅲ), **8b.2**］を適用せよ．］

8　$\mathbf{S} \sim W_p(k, \Sigma)$ で $|\Sigma| \neq 0$ とする．$\mathbf{S}$ の c.f. が

$$E\exp(i\sum\sum t_{rs}S_{rs})=\frac{|\sigma^{rs}|^{k/2}}{|\sigma^{rs}-2it_{rs}|^{k/2}}$$

であることを示せ．この結果は，最初に

$$E\exp(i\sum\sum t_{rs}U_rU_s)$$

を求めることにより，$W_p(k, \Sigma)$ の定義から導かれる．ただし，$(U_1, \cdots, U_p) \sim N_p(0, \Sigma)$．

9　確率変数の分割 $\mathbf{U}'=(\mathbf{U}_1' \,\vdots\, \mathbf{U}_2')$ とそれに対応する Σ の分割を考える．

$$\Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

$\mathbf{U}$ に関する独立な観測値を考え，(S_{ij}) を補正済平方和積和とする．$\mathbf{U}_2$ の1次関数 $\mathbf{L}'\mathbf{U}_2$ を選び，$\mathbf{L}'\mathbf{U}_2$ の $\mathbf{U}_1$ の成分にたいする回帰分析を行え．

9.1　回帰による平方和と回帰からの偏差の平方和は（1変量の公式を用いて）

$$\mathbf{L}'\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{L} \quad および \quad \mathbf{L}'(\mathbf{S}_{22}-\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})\mathbf{L}$$

となることを示せ．ただし $\mathbf{S}_{11}, \mathbf{S}_{12}, \mathbf{S}_{22}$ は，Σ におけると同様な，$\mathbf{S}$ の分割である．

9.2　$\mathbf{U}_1$ と $\mathbf{U}_2$ における成分の数を r および s とする．$\Sigma_{12}=\mathbf{0}$ のとき，1変量理論から

$$\mathbf{L}'\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{L} \quad および \quad \mathbf{L}'(\mathbf{S}_{22}-\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})\mathbf{L}$$

は，r と $(n-1-r)$ D.F. の中心 χ^2 分布に従って，独立に分布することを示せ，$\mathbf{L}$ を落とす方法［(ⅱ), **8b.2**］によって，

$$\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} \sim W_s(r, \Sigma), \qquad \mathbf{S}_{22}-\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} \sim W_s(n-1-r, \Sigma)$$

を求め，それらが独立であることを導け．

9.3　$\Sigma_{12}=\mathbf{0}$ のとき，

$$\frac{|\mathbf{S}|}{|\mathbf{S}_{11}||\mathbf{S}_{22}|}=\frac{|\mathbf{S}_{22}-\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}|}{|\mathbf{S}_{22}|}$$

は，独立なベータ変数の積として分布することを示せ．

10　2人の兄弟のそれぞれについて p 個の測定値が n 家族より得られたとする．標本級内相関を最大にするように測定値の最良1次結合を決定せよ．兄弟の測定値間の相関の有意性を調べる検定法を工夫せよ（詳細は Rao (1945a) 参照）．

11　直角座標　$U_i \sim N_p(\mathbf{0}, \Sigma)$，$i=1, \cdots, k$ は互いに独立であるとする．ウィシャート行列を，$\mathbf{S}=\sum \mathbf{U}_r\mathbf{U}_r'$ と定義する．さらに，$\mathbf{T}$ を $\mathbf{S}$ の（下）三角平方根とする．$\mathbf{T}$ の要素 T_{ij} は，Mahalanobis, Bose and Roy (1937) によって，直角座標と呼ばれた．

11.1　(S_{ij}) から (T_{ij}) への変換を考える．変換のヤコビアンは ($\mathbf{S}=\mathbf{T}\mathbf{T}'$ の関係を用いて)，

$$\frac{D(S_{11}, \cdots, S_{pp})}{D(T_{11}, \cdots, T_{pp})}=(2T_{11})(2T_{11}T_{22})\cdots(2T_{11}T_{22}\cdots T_{pp})$$
$$=2^pT_{11}^pT_{22}^{p-1}\cdots T_{pp}$$

であることを示せ．

11.2　$(S^{ij})=(S_{ij})^{-1}$ とすると，

（a）$|\mathbf{S}|^{1/2}=T_{11}T_{22}\cdots T_{pp}$,

（b）$T_{pp}^2=\dfrac{1}{S^{pp}}$,

（c）$T_{pp}^2 \sim \dfrac{1}{\sigma^{pp}}\chi^2(k-p+1)$

[(c) は [(viii), 8b.2] の結果 (a) と同じことに注意せよ].

11.3　$\boldsymbol{\Sigma}=\mathbf{I}$ とする．関係 $\mathbf{S}=\mathbf{TT}'$ より，$T_{p1},\cdots,T_{p,p-1}$ は $U_{p1},\cdots,U_{pk}$ の1次関数，T_{pp}^2 は2次関数であることを確かめよ．さらに，

$$\sum_{r=1}^{k}U_{pr}^2=S_{pp}=T_{p1}^2+\cdots+T_{p,p-1}^2+T_{pp}^2,$$

および，T_{pp}^2 は $(k-p+1)$D.F. の χ^2 であるので，階級 $k-p+1$ の2次形式である．$\boldsymbol{\Sigma}=\mathbf{I}$ であるから，$U_{p1},\cdots,U_{pk}$ は互いに独立な $N_1(0,1)$ 確率変数である．したがってコクランの定理を適用して

$$T_{pi}^2\sim\chi^2(1),\ \text{つまり},\ T_{pi}\sim N_1(0,1),\quad i=1,\cdots,p-1$$

および $T_{pp}^2\sim\chi^2(k-p+1)$ であり，それらはすべて互いに独立であることがわかる．そこで，すべての T_{ij} は互いに独立に分布し，

$$T_{ij}\sim N_1(0,1),\quad i\neq j\quad \text{および}\quad T_{ii}^2\sim\chi^2(k-i+1)$$

であることを証明せよ．

11.4　したがって，$\boldsymbol{\Sigma}=\mathbf{I}$ のとき，T_{ij} の同時確率密度は，

$$\left[2^{kp/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{k-i+1}{2}\right)\right]^{-1}\tag{A}$$

$$\times e^{-[T_{11}^2+(T_{21}^2+T_{22}^2)+\cdots+(T_{p1}^2+\cdots+T_{pp}^2)]/2}\tag{B}$$

$$\times 2^pT_{11}^{k-1}T_{22}^{k-2}\cdots T_{pp}^{k-p}\tag{C}$$

である．

11.5　$\boldsymbol{\Sigma}=\mathbf{I}$ のとき，(S_{ij}) の同時確率密度は，**11.4** の確率密度を **11.1** のヤコビアンで除し，その式を (S_{ij}) を用いて書くことにより得られる．こうして，(S_{ij}) の確率密度，

$$Ae^{-(S_{11}+S_{22}+\cdots+S_{pp})/2}|\mathbf{S}|^{(k-p-1)/2}$$

が得られる．ここで，定数 A は **11.4** におけると同様である．

11.6　一般の $\boldsymbol{\Sigma}$ にたいする (S_{ij}) の同時確率密度は，変換

$$\mathbf{S}=\mathbf{CS}^*\mathbf{C}',\qquad \mathbf{CC}'=\boldsymbol{\Sigma}$$

をつくることによって得られる．ただし，$\mathbf{S}^*$ は **11.5** の確率密度をもつ．以下を確かめよ．

(a)　$\dfrac{D(\mathbf{S})}{D(\mathbf{S}^*)}=|\mathbf{C}|^{p+1}$　(ヤコビアン)

(b)　$|\mathbf{S}|=|\mathbf{S}^*||\boldsymbol{\Sigma}|$

(c)　$\sum \mathbf{S}_{ii}^*=\mathrm{trace}\mathbf{S}^*=\mathrm{trace}(\mathbf{C}^{-1}\mathbf{SC}'^{-1})=\mathrm{trace}(\mathbf{C}'^{-1}\mathbf{C}^{-1}\mathbf{S})$

$$=\mathrm{trace}\boldsymbol{\Sigma}^{-1}\mathbf{S}=\sum\sum\sigma^{ij}S_{ij}.$$

これより，求める確率密度は

$$A|\boldsymbol{\Sigma}|^{-k/2}|\mathbf{S}|^{(k-p-1)/2}e^{-\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})/2}$$

であり，これは $k\geqslant p$ のときウィシャート分布 $W_p(k,\boldsymbol{\Sigma})$ の確率密度であることを証明せよ．$k<p$ のとき，密度関数は存在しない．

11.7　以上より，または問題 **11.4** より直接に，一般の $\boldsymbol{\Sigma}$ にたいする直角座標の確率密度は

$$2^pA|\boldsymbol{\Sigma}|^{-k/2}T_{11}^{k-1}T_{22}^{k-2}\cdots T_{pp}^{k-p}e^{-\mathrm{trace}(\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T})/2}$$

であることを示せ．

11.8　$\mathbf{U}_1,\cdots,\mathbf{U}_n$ の確率密度関数 $P(\mathbf{U}|\boldsymbol{\Sigma})=P(\mathbf{U}_1|\boldsymbol{\Sigma})\cdots P(\mathbf{U}_n|\boldsymbol{\Sigma})$ は $\mathbf{S}$ と $\boldsymbol{\Sigma}$ しか含まないことを確かめよ．さらに，(この結果は Malay Ghosh による)

$$P(\mathbf{S}|\boldsymbol{\Sigma})=\frac{P(\mathbf{U}|\boldsymbol{\Sigma})}{P(\mathbf{U}|\mathbf{I})}P(\mathbf{S}|\mathbf{I})$$

を示せ．ただし，$P(\mathbf{S}|\boldsymbol{\Sigma})$ は一般の $\boldsymbol{\Sigma}$ にたいする $\mathbf{S}$ の確率密度である ($P(\mathbf{S}|\mathbf{I})$ は **11.4** で導かれているから，$P(\mathbf{S}|\boldsymbol{\Sigma})$ の表現は **11.6** に含まれる難かしい代数演算なしに，上の式を用いて，容易に書き下すことができる)．

[ヒント：$\mathbf{U}$ を $\mathbf{S}$, $\mathbf{Z}$ に変換し，そのヤコビアンを $J(\mathbf{S},\mathbf{Z})$ としよう．そうすれば，

$$P(\mathbf{S}|\boldsymbol{\Sigma})=\int P(\mathbf{U}|\boldsymbol{\Sigma})J(\mathbf{S},\mathbf{Z})d\mathbf{Z}=P(\mathbf{U}|\boldsymbol{\Sigma})\int J(\mathbf{S},\mathbf{Z})d\mathbf{Z}.$$

となり，$\boldsymbol{\Sigma}=\mathbf{I}$ とおくと，

$$P(\mathbf{S}|\mathbf{I})=P(\mathbf{U}|\mathbf{I})\int J(\mathbf{S},\mathbf{Z})d\mathbf{Z}$$

となる．したがって $P(\mathbf{S}|\boldsymbol{\Sigma})/P(\mathbf{S}|\mathbf{I})=P(\mathbf{U}|\boldsymbol{\Sigma})/P(\mathbf{U}|\mathbf{I})$，これはまた $\boldsymbol{\Sigma}$ にたいする統計量 $\mathbf{S}$ の十分性からも証明される]．

12　$\overline{\mathbf{U}}$ を，$N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ から得られた互いに独立な n 個の標本 $\mathbf{U}_1,\cdots,\mathbf{U}_n$ の標本平均，$(\mathbf{S}_{ij})$ を補正済 S.P. 行列とする．このとき，ホテリングの T^2 は，

$$T^2=n(\overline{\mathbf{U}}-\boldsymbol{\mu})(s_{ij})^{-1}(\overline{\mathbf{U}}-\boldsymbol{\mu}),\qquad s_{ij}=S_{ij}/(n-1)$$

によって定義され，[(xii), 8b.2] で示したように

$$\frac{n-p}{p}\frac{T^2}{n-1}\sim F(p,n-p)$$

である．

12.1　次を示せ．

$$T^2=n\max_{\mathbf{L}}\frac{[\mathbf{L}'(\overline{\mathbf{U}}-\boldsymbol{\mu})]^2}{\mathbf{L}'(s_{ij})\mathbf{L}}.$$

12.2　上より，F_α をD.F. p と $(n-p)$ の F 分布の上側 α 点とすれば，

$P\{n[\mathbf{L}'(\overline{\mathbf{U}}-\boldsymbol{\mu})]^2\leqslant\mathbf{L}'(s_{ij})\mathbf{L}[(n-1)pF_\alpha/(n-p)]$ すべての $\mathbf{L}$ にたいして$\}=1-\alpha$ であることを導け．次を確かめよ．

$P\{\mathbf{L}'\overline{\mathbf{U}}-c\sqrt{\mathbf{L}'(s_{ij})\mathbf{L}}\leqslant\mathbf{L}'\boldsymbol{\mu}\leqslant\mathbf{L}'\overline{\mathbf{U}}+c\sqrt{\mathbf{L}'(s_{ij})\mathbf{L}}$，すべての $\mathbf{L}$ にたいして$\}\geqslant 1-\alpha$ ただし，$c^2=(n-1)pF_\alpha/n(n-p)$ で，上式は $\boldsymbol{\mu}$ のすべての1次関数にたいして，同時信頼区間を与える (Roy and Bose (1953))．

13　**重相関係数の分布** (Fisher (1928))　$\mathbf{U}_1,\cdots,\mathbf{U}_n$ を $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ より得られた n 個の互いに独立な観測値とし，$(S_{ij})_k$ を $\mathbf{U}$ の最初の k 個の成分にたいする補正済 S.P. 行列とする．p 番目の成分の他の成分との重相関係数は

$$R^2=1-\frac{1}{S_{pp}}\frac{|S_{ij}|_p}{|S_{ij}|_{p-1}}=\frac{R_1^2-R_0^2}{R_1^2}$$

によって推定される．ただし

$$R_0^2=\min_{\alpha,\beta_i}\sum(U_{pr}-\alpha-\beta_1U_{1r}-\cdots-\beta_{p-1}U_{p-1,r})^2=\frac{|S_{ij}|_p}{|S_{ij}|_{p-1}}$$

$$R_1^2=\min_{\beta_i=0,\alpha}\sum(U_{pr}-\alpha-\beta_1U_{1r}-\cdots-\beta_{p-1}U_{p-1,r})^2=S_{pp}.$$

13.1　最小2乗法の第2基本定理 (176頁) より，次式を導け．

$$R_0^2\sim\sigma^2\chi^2(n-p),\qquad R_1^2-R_0^2\sim\sigma^2\chi^2(p-1,\delta^2)$$

ただし，$\delta^2=\sum\sum_{i,j=1}^{p-1}S_{ij}\beta_i\beta_j\div\sigma^2$，この σ^2 は，$U_1,\cdots,U_{p-1}$ が所与のときの U_p の残差分

散である．したがって，U_{ir}，$i=1,\cdots,p-1\,;\,r=1,\cdots,n$ が所与のときの R^2 の条件付き確率密度は，$\chi^2(p-1,\delta^2)/[\chi^2(p-1,\delta^2)+\chi^2(n-p)]$ の密度であり，それは非心ベータ分布である（第3章の問題 **17.1** と **17.2**）．

$$e^{-\delta^2/2}\sum\left(\frac{\delta^2}{2}\right)^r\frac{1}{r!}B\left(R^2\middle|\frac{p-1}{2}+r,\frac{n-p}{2}\right)$$
$$=e^{-\delta^2/2}B\left(R^2\middle|\frac{p-1}{2},\frac{n-p}{2}\right){}_1F_1\left(\frac{n-1}{2},\frac{p-1}{2},\frac{\delta^2R^2}{2}\right),$$

ただし，$B(x|\alpha,\beta)=\Gamma(\alpha+\beta)x^{\alpha-1}(1-x)^{\beta-1}/\Gamma(\alpha)\Gamma(\beta)$．

13.2　R^2 の条件つき分布は δ^2 しか含まないことを確かめよ．一方，$\delta^2\sigma^2\sim(\sum\sum\sigma_{ij}\beta_i\beta_j)\chi^2(n-1)$．$R^2$ の条件つき確率密度に δ^2 の密度を乗じ，δ^2 に関して項別に積分すれば，R^2 の無条件確率密度

$$(1-\rho^2)^{(n-1)/2}B\left(R^2\middle|\frac{p-1}{2},\frac{n-p}{2}\right){}_2F_1\left(\frac{n-1}{2},\frac{n-1}{2},\frac{p-1}{2},\rho^2R^2\right)$$

が得られる．ただし，ρ^2 は母集団重相関係数

$$\rho^2=\sum_{i,j=1}^{p-1}\sum\sigma_{ij}\beta_i\beta_j\div\sigma^2=1-\frac{1}{\sigma_{pp}\sigma^{pp}}$$
$$=1-\frac{1}{\sigma_{pp}}\frac{|\sigma_{ij}|_p}{|\sigma_{ij}|_{p-1}}$$

である．

14　$\mathbf{S}_1\sim\mathbf{W}_p(k_1,\Sigma)$ と $\mathbf{S}_2\sim\mathbf{W}_p(k_2,\Sigma)$ は互いに独立に分布するとする．以下を証明するために，ウィシャート分布の定義または確率変数の行列 $\mathbf{S}_1$ の確率密度関数を用いよ．

14.1　方程式 $|\mathbf{S}_1-\lambda(\mathbf{S}_1+\mathbf{S}_2)|=0$ の根と $(\mathbf{S}_1+\mathbf{S}_2)$ の要素は，互いに独立に分布する．

14.2　特に，$|\mathbf{S}_1|/|\mathbf{S}_1+\mathbf{S}_2|$ は $\mathbf{S}_1+\mathbf{S}_2$ と互いに独立に分布する．

14.3　これより，$\mathbf{S}_i\sim\mathbf{W}_p(k_i,\Sigma)$，$i=1,2,\cdots$ でこれらはすべて互いに独立であり，かつ，$k_i\geqslant p$ とすると，

$$\frac{|\mathbf{S}_1|}{|\mathbf{S}_1+\mathbf{S}_2|},\quad\frac{|\mathbf{S}_1+\mathbf{S}_2|}{|\mathbf{S}_1+\mathbf{S}_2+\mathbf{S}_3|},\quad\cdots$$

はすべて互いに独立に分布することを導け．

15　**8d.3** で，指定された判別関数および，任意の2つの特定の変数の係数の指定された比にたいする検定について考察した．$x_1,\cdots,x_p$ のうちのたとえば $x_1,\cdots,x_s$ という s 個の特定の変数の指定された比 $\rho_1,\cdots,\rho_s$ にたいする検定とは何であるか？

［ヒント：この検定は $p-s+1$ 個の変数

$$\rho_1x_1+\cdots+\rho_sx_s,\qquad x_{s+1},\ \cdots,\ x_p$$

の十分性を調べることと同じである．もし，それらに基づく D^2 を D^2_{p-s+1}，全変数に基づく D^2 を D^2_p とするならば，その分散比は $s-1$ と N_1+N_2-p-1 D.F の

$$F=\frac{N_1+N_2-p-1}{s-1}\frac{N_1N_2(D^2_p-D^2_{p-s+1})}{(N_1+N_2)(N_1+N_2-2)+N_1N_2D^2_{p-s+1}}$$

となる．記号は **8d.3** 参照．］

16　Δ^2_p と Δ^2_q は，それぞれ，p 個の特性に基づく場合と，その部分集合である q 個の特性に基づく場合の，2つの母集団のあいだのマハラノビス汎距離をあらわすとする．(8d.3.7) の検定が仮説 $\Delta^2_p=\Delta^2_q$ についてであるのにたいし，(8d.3.2) の検定は，仮説 $\Delta^2_p=0$ にたいするものであることを示せ．一般的に $\Delta^2_p\geqslant\Delta^2_q$ であることを確かめよ．

17　**複素正規分布**　　複素 p 次元確率ベクトル $\mathbf{Z}=\mathbf{X}_1+i\mathbf{X}_2$ は，すべての複素ベクトル $\mathbf{L}$ にたいして，

$$\mathrm{R.P.}(\mathbf{L}^*\mathbf{Z})\sim N_1(\mathrm{R.P.}(\mathbf{L}^*\boldsymbol{\mu}),\ \mathbf{L}^*\mathbf{QL})$$

ならば，複素 p 変量正規分布に従って分布するといわれる．ここで，R.P. は実数部をあらわし，$\boldsymbol{\mu}$ は複素 p 次元ベクトル，$\mathbf{Q}$ は $p\times p$ エルミート非負定符号行列である．このようなとき，次のように書く．

$$\mathbf{Z}\sim CN_p(\boldsymbol{\mu},\mathbf{Q}).$$

$\boldsymbol{\mu}=\boldsymbol{\mu}_1+i\boldsymbol{\mu}_2$，$\mathbf{Q}=\mathbf{Q}_1+i\mathbf{Q}_2$，$\mathbf{v}'=(\boldsymbol{\mu}_1'\,\vdots\,\boldsymbol{\mu}_2')$ および

$$\Sigma=\begin{pmatrix}\mathbf{Q}_1 & -\mathbf{Q}_2\\ \mathbf{Q}_2 & \mathbf{Q}_1\end{pmatrix}$$

とあらわす．以下を示せ．

（i）$\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}\sim N_{2p}(\mathbf{v},\Sigma)$

（ii）R.P.$(\mathbf{L}^*\mathbf{Z})$ と I.P.$(\mathbf{L}^*\mathbf{Z})$ は，任意の $\mathbf{L}$ にたいして，互いに独立に分布する．

（iii）I.P.$(\mathbf{L}^*\mathbf{Z})\sim N_1(\mathrm{I.P.}(\mathbf{L}^*\boldsymbol{\mu}),\ \mathbf{L}^*\mathbf{QL})$，ただし I.P. は虚数部をあらわす．

参考文献

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Ashton, E. H., M. J. R. Healy and S. Lipton (1957), The descriptive use of discriminant functions in physical anthropology, *Proc. Roy. Soc.* (*series B*) **146**, 552–572.

Barnard, M. M. (1935), The secular variations of skull characters in four series of Egyptian skulls, *Ann. Eugen.* **6**, 352–371.

Bartlett, M. S. (1947a), Multivariate analysis, *J. Roy. Stat. Soc. Supple* **9** (**B**), 176–197.

Bartlett, M. S. (1947b), The general canonical correlation distribution, *Ann. Math. Statist.* **18**, 1–17.

Bartlett, M. S. (1950), Tests of significance in factor analysis, *Brit. J. Psych.* (*Stat. Sec.*) **3**, 77–85.

Bose, R. C. and S. N. Roy (1938), The distribution of Studentised D^2-statistic, *Sankhyā* **4**, 19–38.

Bowker, A. H. (1960), A representation of Hotelling's T^2 and Anderson's classification statistic, *Contributions to Probability and Statistics* (*Essays in honour of Harold Hotelling*), 142–149, Stanford University Press.

Burnaby, T. P. (1966), Growth invariant discriminant functions and generalized distance, *Biometrics* **22**, 96–110.

Cramér, H. (1937), *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Math. Phy. **36**, Cambridge University Press.

Fisher, R. A. (1928), The general sampling distribution of the multiple correlation coefficient, *Proc. Roy. Soc.* **A 121**, 654–673.

Fisher, R. A. (1936), The use of multiple measurements in taxonomic problems, *Ann. Eugen.* **7**, 179-188.

Fisher, R. A. (1939), The sampling distribution of some statistics obtained from nonlinear equations, *Ann. Eugen.* **9**, 238-249.

Fisher, R. A. (1940), The precision of discriminant functions, *Ann. Eugen.* **10**, 422-429.

Fréchet, M. (1951), Generalisation de la loi probabilite de Laplace, *Ann. de l'Institute Henri Poincare* **12**.

Girshick, M. A. (1939), On the sampling theory of roots of determinantal equations, *Ann. Math. Statist.* **10**, 203-224.

Grenander, U. and M. Rosenblatt (1957), *Statistical Analysis of Stationary Time Series*, Wiley, New York, Almqvist and Wicksell, Stockholm.

Hotelling, H. (1931), The generalization of Student's ratio, *Ann. Math. Statist.* **2**, 360-378.

Hotelling, H. (1935), The most predictable criterion, *J. Educ. Psych.* **26**, 139-142.

Hotelling, H. (1936), Relations between two sets of variates, *Biometrika* **28**, 321-377.

Hsu, P. L. (1938), Notes on Hotelling's generalised T^2, *Ann. Math. Statist.* **9**, 231-243.

Hsu, P. L. (1939), On the distribution of the roots of certain determinantal equations, *Ann. Eugen.* **9**, 250-258.

James, A. T. (1954), Normal multivariate analysis and the orthogonal group, *Ann. Math. Statist.* **25**, 40-75.

John, S. (1961), Errors in discrimination, *Ann. Math. Statist.* **32**, 1125-1144.

Khatri, C. G. (1968), Some results for the singular normal multivariate regression models, *Sankhyā* **A 30**, 267-280.

Kshirsagar, A. M. (1961), The goodness of fit of a single (non-isotropic) hypothetical principal component, *Biometrika* **48**, 397-407.

Kshirsagar, A. M. (1972), *Multivariate Analysis*, Marcell Dekker, Inc. New York.

Lawley, D. N. and A. E. Maxwell (1963), *Factor Analysis as a Statistical Method*, Butterworth's Mathematical Texts, London.

Mahalanobis, P. C. (1936), On the generalized distance in statistics, *Proc. Nat. Inst. Sci. India* **12**, 49-55.

Mahalanobis, P. C., R. C. Bose and S. N. Roy (1937), Normalization of variates and the use of rectangular coordinates in the theory of sampling distributions, *Sankhyā* **3**, 1-40.

Mitra, S. K. (1970), Some characteristic and non-characteristic properties of the Wishart distribution, *Sankhyā* **A 31**, 19-22.

Nair, U. S. (1939), The application of the moment function in the study of distribution laws in statistics, *Biometrika* **30**, 274-294

Prohorov, Yu. V. and M. Fisz (1957), A characteristic property of the normal distribution in Hilbert space, *Teoria Veroiatnostei iee Primeneniia* **2**, 475-477.

Potthoff, R. F. and S. N. Roy (1964), A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* **51**, 313-326.

Rao, K. S. (1951), On the mutual independence of a set of Hotelling's T^2 derivable from a sample of size n from a k-variate normal population, *Bull. Inst. Internat. Stat.* **33**, Part II, 171-176.

Rasch, G. (1953), On simultaneous factor analysis in several populations, *Uppsala Symposium on Psychological Factor Analysis, Uppsala*, Almqvist and Wicksell,

65-71.

Riffenburg, R. H. and C. W Clunies-Ross (1960), Linear discriminant analysis, *Pacific Science* **14**, 251-256.

Roy, J. (1951), The distribution of certain likelihood criteria useful in multivariate analysis, *Bull. Inst. Internat. Stat.* **33**, Part II, 219-230.

Roy, S. N. (1939), p-Statistics or some generalizations in the analysis of variance appropriate to multivariate problems, *Sankhyā* **4**, 381-396.

Roy, S. N. (1957), *Some Aspects of Multivariate Analysis*, Wiley, New York.

Roy, S. N and R. C. Bose (1953), Simultaneous confidence interval estimation, *Ann. Math. Statist.* **24**, 513-536.

Roy, S. N., R. Gnanadesikan and J. N. Srivastava (1972), *Analysis and Design of Certain Quantitative Multiresponse Experiments*, Pergamon Press.

Smith, C. A. B. (1947), Some examples of discrimination, *Ann. Eugen.* **13**, 272-282.

Wald, A. (1944), On a statistical problem arising in the classification of an individual into one of two groups. *Ann. Math. Statist.* **15**, 145-162.

Watson, G. W. (1964), A note on maximum likelihood, *Sankhyā* **26 A**, 303-304.

Wijsman, R. A. (1957), Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis, *Ann. Math. Statist.* **28**, 415-423.

Wilks, S. S. (1932), Certain generalizations in the analysis of variance, *Biometrika* **24**, 471-494.

Wilks, S. S. (1962), *Mathematical Statistics*, Wiley, NewYork.

Williams, E. J. (1952), Some exact tests in multivariate analysis, *Biometrika* **39**, 17-31.

Williams, E. J. (1955), Significance tests for discriminant functions and linear functional relationships, *Biometrika* **42**, 360-381.

Williams, E. J. (1959), *Regression Analysis*, Wiley, New York.

Williams, E. J. (1967), The analysis of association among many variates. *J. Roy Statist. Soc.* **B 29**, 199-228.

Wishart, John (1928), The generalized product moment distribution in samples from a normal multivariate population, *Biometrika* **20 A**, 32-52.

Wold, H. (1938), *A Study in the Analysis of Stationary Time Series*, University of Uppsala

著者の刊行物

書　籍

B1. *Linear Statistical Inference and Its Applications.* John Wiley, 1965, 1973 (second edition).

B2. *Advanced Statistical Methods in Biometric Research* John Wiley, 1952 and Heffner, 1971

B3. (S.K. Mitra と共著) *Generalized Inverse of Matrices and Its Applications.* John Wiley, 1971.

B4. (A. Kagan and Yu. V. Linnik と共著) *Characterization Problems of Mathematical Statistics* (ロシア語) U.S.S.R Academy of Sciences, 1972. (英訳は John Wiley, 印刷中).

B5. *Computers and the Future of Human Society.* Andhra University and Statistical Publishing Society, 1970

B6. (R.K. Mukherjee and J.C. Trevor と共著) *The Ancient Inhabitants of Jebel Moya* Cambridge University Press, 1955.

B7. (P.C. Mahalanobis and D.N. Majumdar と共著) *Anthropometric Survey of the United Provinces. 1941.* A statistical study. *Sankhyā* **9**, 90–324, 1949.

B8. (Majumdar と共著) *Bengal Anthropometric Survey, 1945.* A statistical study. *Sankhyā* **19**, 201–408；また，次の標題の本に納められている. *Race Elements of Bengal, a Quantitative Study.* Asia Publishing House, 1958.

B9. (A. Matthai and S.K. Mitra と共著) *Formulae and Tables for Statistical Work.* Statistical Publishing Society, 1966

研 究 論 文

1941

a. (K.R. Nair と共著) Confounded designs for asymmetrical factorial experiments. *Science and Culture* **7**, 361.

1942

a. On the volume of a prismoid in N-space and some problems in continuous probability. *Math. Student* **10**, 68–74.

b. On the sum of observations from different Gamma type populations. *Science and Culture* **7**, 614.

c. (K.R. Nair と共著) Confounded designs for $k\times p^m\times q^m\times\cdots$ type of factorial experiments. *Science and Culture* **7**, 361.

d. (K.R. Nair と共著) A general class of quasi-factorial designs leading to confounded designs for factorial experiments. *Science and Culture* **7**, 457.

e. (K.R. Nair と共著) A note on partially balanced incomplete block designs. *Science and Culture* **7**, 568.

f. (K.R. Nair と共著) Incomplete block designs involving several groups of varieties. *Science and Culture* **7**, 625.

1943

a. Certain experimental arrangements in quasi-latin squares. *Current Science* **12**, 322.
b. On bivariate correlation surfaces. *Science and Culture* **8**, 236.

1944

a. (R.C. Bose and S. Chowla と共著) On the integral order mod p of quadratics x^2+ax+b with applications to the construction of minimum functions for $GF(p^2)$ and to some number theory results. *Bull. Cal. Math. Soc.* **36**, 153-174.

1945

a. Familial correlations or the multivariate generalization of the intraclass correlation. *Current Science* **14**, 66.
b. Generalization of Markoff's theorem and tests of linear hypotheses. *Sankhyā* **7**, 9-16.
c. Markoff's theorem with linear restrictions on parameters. *Sankhyā* **7**, 16-19.
d. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.* **37**, 81-91.
e. Finite geometries and certain derived results in theory of numbers. *Proc Nat. Inst. Sc.* **11**, 136-149.
f. (R.C. Bose and S. Chowla と共著) On the roots of a well-known congruence. *Proc. Nat. Acad. Sc.* **14**, 193.
g. (R.C. Bose and S. Chowla と共著) Minimum functions in Galois fields. *Proc. Nat. Acad. Sc.* **14**, 191.

1946

a. Difference sets and combinatorial arrangements derivable from finite geometries *Proc. Nat. Inst. Sc.* **12**, 123-135.
b. On the linear combination of observations and the general theory of least squares. *Sankhyā* **7**, 237-256.
c. Confounded factorial designs in quasi-latin squares. *Sankhyā* **7**, 295-304.
d. Hypercubes of strength 'd' leading to confounded designs in factorial experiments. *Bull. Cal Math. Soc.* **38**, 67-78.
e. On the mean conserving property. *Proc. Ind. Acad. Sc.* **23**, 165-173.
f. Tests with discriminant functions in multivariate analysis. *Sankhyā* **7**, 407-414.
g. On the most efficient designs in weighing. *Sankhyā* **7**, 440.
h. Minimum variance and the estimation of several parameters. *Proc. Camb. Phil. Soc.* **43**, 280-283.
i. (S.J Poti と共著) On locally most powerful tests when alternatives are one sided. *Sankhyā* **7**, 441.

1947

a. The problem of classification and distance between two populations. *Nature* **159**, 30.
b. Note on a problem of Ragnar Frisch. *Econometrika* **15**, 245-249. A correction to Note on a problem of Ragnar Frisch. *Econometrika* **17**, 212.
c. General methods of analysis for incomplete block designs. *J. Am. Statist. Assoc.* **42**, 541-561

d. Factorial experiments derivable from combinatorial arrangements of arrays. *J. Roy. Statist. Soc.* **B9**, 128-140.
e. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Camb. Phil. Soc.* **44**, 50-57.

1948

a. A statistical criterion to determine the group to which an individual belongs. *Nature* **160**, 835.
b. Tests of significance in multivariate analysis. *Biometrika* **35**, 58-79.
c. The utilization of multiple measurements in problems of biological classification. *J. Roy. Statist. Soc.* **B10**, 159-203.
d. Sufficient statistics and minimum variance estimates. *Proc. Camb. Phil. Soc.* **45**, 215-218.
e. (D.C. Shaw と共著) On a formula for the prediction of cranial capacity. *Biometrics* **4**, 247-253.
f. (K.R. Nair と共著) Confounding in asymmetrical factorial experiments. *J. Roy. Statist. Soc.* **B10**, 109-131.

1949

a. On a class of arrangements. *Edin. Math. Proc.* **8**, 119-125
b. On some problems arising out of discrimination with multiple characters. *Sankhyā* **9**, 343-364.
c. A note on unbiased and minimum variance estimates. *Cal. Statist. Bull.* **3**, 36.
d. (P. Slater と共著) Multivariate analysis applied to differences between neurotic groups. *British J Psychology, Statist. Sec.* **2**, 17-29.

1950

a. Methods of scoring linkage data giving the simultaneous segregation of three factors. *Heredity* **4**, 37-59.
b. The theory of fractional replication in factorial experiments. *Sankhyā* **10**, 84-86.
c. Statistical inference applied to classificatory problems. *Sankhyā* **10**, 229-256.
d. A note on the distribution of $D^2_{p+q}-D^2_p$ and some computational aspects of D^2 statistic and discriminant function. *Sankhyā* **10**, 257-268.
e. Sequential tests of null hypotheses. *Sankhyā* **10**, 361-370.

1951

a. A theorem in least squares. *Sankhyā* **11**, 9-12.
b. Statistical inference applied to classificatory problems. Part II. Problems of selecting individuals for various duties in a specified ratio. *Sankhyā* **11**, 107-116.
c. A simplified approach to factorial experiments and the punched card technique in the construction and analysis of designs. *Bull. Inst. Inter. Statist.* **XXXIII(2)**, 1-28.
d. An asymptotic expansion of the distribution of Wilks' criterion. *Bull Inst. Inter. Statist.* **XXXIII(2)**, 177-180.
e. The applicability of large sample tests for moving average and auto regressive schemes to series of short length—an experimental study. Part 3: The discriminant function approach in the classification of time series. *Sankhyā* **11**, 257-272.

f. Progress of statistics in India. *Progress of Science in India* (1939-1950), 68-94, National Institute of Sciences. India.

1952

a. Some theorems on minimum variance estimation. *Sankhyā* **12**, 27-42.
b. Minimum variance estimation in distributions admitting ancillary statistics *Sankhyā* **12**, 53-56.

1953

a. Discriminant function for genetic differentiation and selection. *Sankhyā* **12**, 229-246.
b. On transformations useful in the distribution problems of least squares. *Sankhyā* **12**, 339-346.

1954

a. A general theory of discrimination when the information about alternative population distribution is based on samples. *Ann. Math Statist.* **25**, 651-670.
b. On the use and interpretation of distance functions in statistics. *Bull Inst. Inter. Statist.* **34**, 90-100.
c. Estimation of relative potency from multiple response data. *Biometrics* **10**, 208-220.

1955

a. Analysis of dispersion for multiply classified data with unequal numbers in cells. *Sankhyā* **15**, 253-280
b. Estimation and tests of significance in factor analysis. *Psychometrika* **20**, 93-111.
c. Theory of the method of estimation by minimum chi-square. *Bull. Inst. Inter. Statist.* **35**, 25-32.
d. (G. Kalianpur と共著) On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhyā* **16**, 331-342.

1956

a. On the recovery of interblock information in varietal trials. *Sankhyā* **17**, 105-114.
b. A general class of quasi-factorial and related designs. *Sankhyā* **17**, 165-174.
c. Analysis of dispersion with missing observations. *J. Roy. Statist. Soc.* **B18**, 259-264.
d. (I. M. Chakravarti と共著) Some small sample tests of significance for a Poisson distribution. *Biometrics* **12**, 264-282.

1957

a. Maximum likelihood estimation for multinomial distribution *Sankhyā* **18**, 139-148

1958

a. Quantitative studies in Sociology, need for increased use in India. *Sociology, Social Research, and Social Problems in India.* Asia Publishing House, 1961, 53-74.
b. Some statistical methods for comparison of growth curves. *Biometrics* **14**, 1-17.

c. Maximum likelihood estimation for the multinomial distribution with an infinite number of cells. *Sankhyā* **20**, 211-218.

1959

a. Some problems involving linear hypotheses in multivariate analysis. *Biometrika* **46**, 49-58
b. Expected values of mean squares in the analysis of incomplete block experiments and some comments based on them. *Sankhyā* **21**, 327-336.
c. Sur une Characterisation de la Distribution Normal Etablie d'apres une Propriete Optimum des Estimations Lineares. *Coll. Inter. due C. N. R. S.* France, **LXXXVII**, 165.
d. (I. M. Chakravarthy と共著) Tables for some small sample tests of significance for Poisson distributions and 2×3 contingency tables. *Sankhyā* **21**, 315-326.

1960

a. Multivariate analysis: An indispensable statistical aid in applied research. *Sankhyā* **22**, 317-338.
b. Experimental designs with restricted randomization. *Bull. Inst. Inter. Statist.* **XXXVII**, 397-404.

1961

a. A study of large sample test criteria through properties of efficient estimates. *Sankhyā A* **23**, 25-40.
b. Asymptotic efficiency and limiting information. *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California, **1**, 531-546.
c. A study of BIB designs with replications 11 to 16. *Sankhyā* **23**, 117-127.
d. A combinatorial assignment problem. *Nature* **191**, 100.
e. Generation of random permutations of given number of elements using random sampling numbers. *Sankhyā A* **23**, 305-307.
f. Combinatorial arrangements analogous to orthogonal arrays. *Sankhyā A* **23**, 283-286.
g. Some observations on multivariate statistical methods in anthropological research. *Bull. Inst. Inter. Statist.* **XXXVII**(4), 99-109.

1962

a. Ronald Aylmer Fisher, F. R. S. (an obituary) *Science and Culture* 29, 80-81.
b. Some observations on anthropometric surveys: *Anthropology today Essays in memory of D. N Majumdar*, 135-149. Asia Publishing House.
c. Use of discriminant and allied functions in multivariate analysis. *Sankhyā A* **24**, 149-154
d. Efficient estimates and optimum inference procedures in large samples (with discussion). *J. Roy. Statist. Soc.* **B 24**, 46-72.
e. Problems of selection with restriction. *J Roy. Statist. Soc.* **B 24**, 401-405.
f. A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *J Roy. Statist Soc.* **B 24**, 152-158.
g. Apparent anomalies and irregularities in maximum likelihood estimation (with

discussion), *Sankhyā* *B* **24**, 73–102.

1963

a. Criteria of estimation in large sampes. *Sankhyā* *A* **25**, 189–206.
b. (V. S. Varadarajan と共著) Discrimination of Gaussian processes. *Sankhyā* *A* **25**, 303–330

1964

a. Problems of selection involving programming techniques. *Proceedings of the IBM Scientific Computing Symposium on Statistics*, 29–51.
b. The use and interpretation of principal component. analysis in applied research. *Sankhyā* *A* **26**, 329–358.
c. (H. Rubin と共著) On a characterization of the Poisson distribution. *Sankhyā* *A* **26**, 295–298.

1965

a. Efficiency of an estimator and Fisher's lower-bound to asymptotic variance. *Bull. Inst. Inter. Statistic* XLI(1), 55–63.
b. The theory of least squares when the parameters are stochastic and its applications to the analysis of growth curves. *Biometrika* **52**, 447–458.
c. On discrete distributions arising out of methods of ascertainment. *Sankhyā* *A* **27**, 311–324.
d. Covariance adjustment and related problems in multivariate analysis. Dayton Symposium on Multivariate Analysis. *Multivariate Analysis* **1**, 87–103. Academic Press Inc., New York.
e. (A. M. Kagan and Yu. V. Linnik と共著) On a characterization of the normal law based on a property of the sample average. *Sankhyā* *A* **27**, 405–406.

1966

a. Characterization of the distribution of random variables in linear structural relations. *Sankhyā* *A* **28**, 251–260.
b. Generalized inverse for matrices and its applications in mathematical statistics —*Festschrift for J Neyman. Research Papers in Statistics*. 263–299.
c. Discriminant function between composite hypotheses and related problems. *Biometrika* **53**, 315–321.
d. Discrimination among groups and assigning new individuals. *The Role of Methodology of Classification in Psychiatry and Psychopathology*, 229–240, U. S. Department of Health, Education, and Welfare, Public Health Services.
e. On some characterisations of the normal law. *Sankhyā* *A* **29**, 1–14.
f. (M. N. Rao と共著) Linked cross sectional study for determining norms and growth curves: A pilot survey on Indian school-going boys. *Sankhyā* *B* **28**, 231–252 (一部は次の標題で刊行されている：Methods for determining norms and growth rates, A study amongst Indian school-going boys. *Gerentologia* **12**, 200–216).

1967

a. Calculus of generalized inverses of matrices: Part I—general theory. *Sankhyā* *A* **29**, 317–350.
b. Cyclical generation of linear subspaces in finite geometries, 515–535, Chapter 28 in *Combinatorial Mathematics and Its Applications*. The University of North Carolina Monograph Series No. 4.

c. Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 355–372, University of California Press.
d. On vector variables with a linear structure and a characterization of the multivariate normal distribution. *Bull Inst. Inter. Statist.* **XLII(2)**, 1207–1213.

1968

a. A note on a previous lemma in the theory of least squares and some further results. *Sankhyā* *A* **30**, 259–266.
b. (C. G. Khatri と共著) Some characterizations of the gamma distribution. *Sankhyā* *A* **30**, 157–166.
c. (C. G. Khatri と共著) Solutions to some functional equations and their applications to characterisation of probability distributions. *Sankhyā* *A* **30**, 167–180.
d. (B. Ramachandran と共著) Some results on characteristic functions and characterizations of the normal and generalized stable laws. *Sankhyā* *A* **30**, 125–140.
e. (S. K. Mitra と共著) Simultaneous reduction of a pair of quadratic forms. *Sankhyā* *A* **30**, 313–322.
f. (S. K. Mitra と共著) Some results in estimation and test of linear hypotheses under the Gauss-Markoff model. *Sankhyā* *A* **30**, 281–290.

1969

a. A decomposition theorem for vector variables with a linear structure. *Ann. Math. Statist.* **40**, 1845–1849.
b. Some characterizations of the multivariate normal distribution. *Multivariate Analysis* **2**, 322–328. Academic Press Inc., New York.
c. Recent advances in discriminatory analysis. *J. Ind. Soc. Agri. Statist.* **XXI**, 3–15.
d. A multidisciplinary approach for teaching statistics and probability. *Sankhyā* *B* **31**, 321–340.
e. (S. K. Mitra と共著) Conditions for optimality and validity of least squares theory. *Ann. Math. Statist.* **40**, 1716–1724.

1970

a. Estimation of heteroscedastic variances in a linear model. *J. Am. Statist. Assoc.* **65**, 161–172.
b. Computers: A great revolution in scientific research. *Proc. Indian National Science Academy* **36**, 123–139.
c. Inference on discriminant function coefficients. *Essays in Probability and Statistics*. Chapter 30, 587–602. Edited by R. C. Bose and others. University of North Carolina and Statistical Publis ing Society.
d. (B. Ramachandran と共著) Solutions of functional equations arising in some regression problems and a characterisation of the Cauchy law. *Sankhyā* *A* **32**, 1–30.

1971

a. Characterisation of probability laws through linear functions *Sankhyā A* 33, 265–270.

b. Some aspects of statistical inference in problems of sampling from finite populations. *Foundations of Statistical Inference*, Holt, Rinehart and Winston of Canada, 177–202.

c. Estimation of variance and covariance components—Minque Theory. *J. Multivariate Analysis* **1**, 257–275.

d. Minimum variance quadratic unbiased estimation of variance components. *J. Multivariate Analysis* **1**, 445–456.

e. Unified theory of linear estimation. *Sankhyā A* 33, 370–396 and *Sankhyā A* 34, 477.

f. Data, analysis, and statistical thinking. *Economic and Social Development, Essays in Honour of C. D. Deshmukh*, Vora and Co. Bombay, 383–392.

g. (J. K. Ghosh と共著) A note on some translation—parameter families of densities for which the median is an m. l. e. *Sankhyā A* 33, 91–93.

h. (S. K. Mitra と共著) Further contributions to the theory of generalized inverse of matrices and its applications. *Sankhyā A* 33, 289–300.

i. (S. K. Mitra と共著) Generalized inverse of a matrix and applications. *Sixth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 601–620, University of California Press.

j. Some comments on the logarithmic seriesdistribution. *Statistical Ecology* **1**, 131–142 Pennsylvania State University Press.

1972

a. Recent trends of research work in multivariate analysis. *Biometrics* **28**, 3–22

b. Estimation of variance and covariance components in linear models. *J. Am. Statist. Assoc.* 67, 112–115.

c. A note on IPM method in the unified theory of linear estimation. *Sankhyā A* 34, 285–288.

d. Some recent results in linear estimation. *Sankhyā*, **B** 34, 369–377.

e. (P. Bhimasankaram and S. K. Mitra と共著) Determination of a matrix by its subclasses of *g*-inverse. *Sankhyā A* 34, 5–8.

f. (C. G. Khatri と共著) Functional equations and characterisation of probability laws through linear functions of random variables. *J. Multivariate Analysis* **2**, 162–173.

1973

a. Unified theory of least squares. *Communication in Statistics*, **1**, 1–8.

b. Some combinatorial problems of arrays and applications to design of experiments, in *A Survey of Combinatorial Theory*, North Holland, Chapter 29, pp. 349–360.

索　　引

ア　行

一般逆行列→逆行列の計算を見よ
1次方程式　linear equations　6
イェーツの連続性補正　Yates' continuity correction　377, 378
因子分析　factor analysis　531
ウィシャート分布　Wishart distribution　482, 540
ウイルクスのΛ検定基準　Wilks' Λ criterion
　定義　499
　——の分布（正確な）distribution　503
　——の分布（漸近的）　503
　バートレットの近似
　　Bartlett's approximation　503
　分解　decomposition　501
　分散比近似（ラオ）
　　variance ratio approximation　503
エントロピー　entropy　160

カ　行

ガウス＝マルコフの理論（1変量）
　ガウス＝マルコフ模型
　　Gauss-Markoff model　201
　仮説の検定　tests of hypotheses　217–219
　計画行列の選択
　　choice of design matrix　215
　推定量の分散
　　variance of estimator　206, 211
　BLE（最良線形）　best linear　278
　BLUE（最良線形不偏）
　　best linear unbiased　205
　観測値の分散共分散行列が特異→線形推定の統一理論を見よ
　幾何学的解　geometric solution　209
　最小2乗法の理論
　　least squares theory　203
　推定量の有効性
　　efficiency of estimator　213
　正規方程式　normal equation　203, 211
　相関のある観測値
　　correlated observation　210
　同時信頼区間　simultaneous confidence interval　219
　同時推定　simultaneous estimaiton　213
　母数が確率変数　random parameter　214
　BLIMBE（最良線形偏り最小）
　　best linear minimum bias　279
　母数が確率変数　random parameter　214
　母数に制約のあるとき
　　restricted parameter　212
ガウス＝マルコフの理論（多変量）
　一般模型　model (general)　503, 511
　応用　509–518
　共分散分析　analysis of dispersion　500
　構造関係（次元数）　structural relationship (dimensionality)　503, 511
　推定（BLUE など）　estimation　493–495
　成長モデル　growth moodel　507
　線形仮説の検定
　　tests of linear hypotheses　500
　付加情報の検定
　　test for additional information　499
仮説検定（理論）　testing of hypotheses
　一様最強力検定
　　uniformly most powerful test　407
　確率化検定　randomized test　408
　確率比検定　probability ratio test　410
　頑健性　robustness　451
　棄却域　critical region　404
　局所最強力検定
　　locally most powerful test　411
　検定関数　critical function　408
　漸近効率　asymptotic efficiency　421
　相似検定　similar region test　414
　単調尤度比　monotone likelihood ratio　418
　並べかえ検定　permutation test　455
　2種類の誤り　two kinds of errors　404
　ネイマン＝ピアソンの補助定理　405–406
　ピットマンの効率　Pitman efficiency　423

フィッシャー＝ベーレン問題
Fisher-Behren's problem 420
不偏検定 unbiased test 421
複合仮説 composite hypothesis 404, 413
分布によらない方法
distribution free methods 446
補助情報 ancillary information 457
有意水準 level of significance 404
リンドレーの逆説 Lindley's paradox 462
回帰（例） regression
係数の一様性 equality of coefficients 256
指定された――の検定
specification (test of) 258
頭蓋容量 cranial capacity 247
回帰と関連性 regression and association
因子分析 factor analysis 531
級内相関 intraclass correlation 245
共通因子 common factors 529
共分散分析 covariance analysis 263
最良線形予測子 best linear predictor 243
最良予測子 best predictor 241
重相関 multiple correlation 243, 245
正準相関 canonical correlation 527
正準変数 canonical variable 528
制約つき係数 restricted coefficients 261
線形予測子 linear predictor 243, 244
相関比 correlation ratio 242
――の一般理論 general theory of 241-247
部分相関 part correlation 283
偏重相関 partial multiple correlation 245
偏相関係数
partial correlation coefficient 245
偏相関比 partial correlation ratio 245
階乗積率 factorial moments 95
確率 probability
――空間 space 78
公理 axioms 76
事後―― posterior 306
事前―― prior 306
条件つき―― conditional 83
――測度 measure 77
――測度の族 family of measure 123
――の独立性 independence 84
――母関数 generating function 94
――密度 density 82, 83
確率変数 random variables
コルモゴロフの一致定理
Kolmogorov consistency theorem 100
条件つき分布 conditional distribution 85
定義 79, 81, 83
――の独立性 independence 85
確率変数の直交基底 orthogonal basis of
a random variable 531
確率母関数
probability generating functions
定義 94
2項分布の―― binomial 94
――の収束 convergence 119
ポアソン分布の―― Poisson 95
管理図（の動作特性） control charts
(O. C. of) 461, 462
頑健性 robustness 450
キュミュラント cumulant 94
棄却差 critical difference 226
期待値 expectation
条件つき――（統計量を与えたときの）
conditional (given statistic) 89
条件つき――の代用 (substitution in) 142
条件つき――(部分集合を与えたときの)
(given subfield) 133
積率 moments 87
定義と定理 86
棄却域 critical region 404
逆行列の計算 matrix inversion
一般逆行列 g-inverse 23
一般逆行列の計算
computation of g-inverse 25, 32, 205
$\mathbf{A}_m^-$ と $\mathbf{A}_l^-$ の間の双対関係
duality between $\mathbf{A}_m^-$ and $\mathbf{A}^-$ 47
最小2乗型逆行列 least square inverse 46
ノルム最小型一般逆行列
minimum norm inverse 45
ノルム最小の最小2乗型一般逆行列
minimum norm least square inverse 46
反射型一般逆行列 reflexive g-inverse 24
左逆行列 left inverse 16
右逆行列 right inverse 16
ムーア＝ペンローズ逆行列
Moore-Penrose inverse 25, 47
共分散分析 analysis of covariance
1変量 univariate 263
多変量 multivariate 499, 507

補助変数 concomitant variables 262
例題 264
極限定理 → 中心極限定理, 大数の法則, 漸近分布, 確率変数の収束, 分布関数の収束を見よ
行列 matrix
一般逆行列 → 逆行列の計算を見よ
エルミート―― hermitian 16, 39
基本―― elementary 16
逆―― inverse 15
共役転置 conjugate transpose 15
グラム―― gramian 65
ケーレイ＝ハミルトンの定理 42
固有値とベクトル
eigen values and vectors 36
巡回―― circulant 64
随伴―― adjoint 71
正規―― normal 40
対称―― symmetric 16
単位―― identity (unit) 14
直交―― orthogonal 16
転置―― transpose 15
トレース（対角和） trace 27, 31, 61
――の階数 rank 15
――の導関数 derivative 67
非負―― non-negative 42
分割―― partitioned 16
ベキ等―― idempotent 27
ユニタリ―― unitary 16
零―― null 14
零化数 nullity 26
零空間 null space 26, 30
列空間 range space 26
行列式 determinants 21, 30
行列の積 matrix product
アダマール積 28
新しい積（カトリ＝ラオ） new product 29
クロネッカー積 28
行列の標準化 matrix reduction
エルミート行列 Hermitian matrix 39
エルミート正準形
Hermite canonical form 18
階数分解 rank factorization 18
QR 分解 Q-R-decomposition 20
極形式標準化 polar form 66
行列の対 pairs of matrices 38
交換可能行列 commuting matrix 39
三角形 triangular form 19, 20
ジョルダン形 Jordan form 42
スペクトル分解
spectral decomposition 38, 42
正規行列 normal matrix 40
対角形 diagonal form 18, 19
対称行列 symmetric matrix 19, 37
台形 echelon form 17
特異値分解
singular value decompositon 40
ハウスホルダー変換
Householder transformation 20
非対称行列 nonsymmetric matrix 40
非負行列 non-negative matrix 42
フロベニウスの定理 Frobenius 43
ペロンの定理 Perron's theroem 43
メツラー行列 Metzler matrix 43
グラム＝シュミットの直交化
Gram-Schmidt orthogonalization 9
経験分布関数
empirical distribution function
コルモゴロフ＝スミルノフ統計量の――
Kolmogorov-Smirnov statistic 384
コルモゴロフの定理 384
収束（グリベンコ） convergence 383
スミルノフの定理 384
定義 382
継続の法則 law of succession 308
決定理論 decision theory
確率化方式 randomized rule 444
完全類（最小）
complete class (minimal) 448
許容的決定方式 admissible rules 445
非確率化方式 non-randomized rule 444
ベイズ解 Bayes' solution 445
ミニマックス方式 minimax rule 449
固有値 eigen values 36
固有値（制限つき） 47
固有方程式 determinantal equation 63

サ　行

最小2乗法 → ガウス＝マルコフ, 線形推定を見よ
最大尤度 maximum likelihood (m. l.)
最尤推定にたいする得点法
scoring for m. l. estimation 336
最尤 (m. l.) 推定量 m. l. estimator 324

最尤推定量による特徴づけ
characterization by m. l. estimators 346
最尤推定量の性質
property of m. l. estimator 334
最尤推定量の分布
distribution of m. l. estimators 334
最尤に近い推定量
near m. l. estimator 323
最尤方程式推定量
m. l. equation estimator 323
最尤法の限界
limitation of m. l. method 325
多項分布の最尤推定 m. l. estimation of multinomial 326-333
シルベスターの法則（行列の階数）
Sylvester's law (matrix rank) 29
指数型分布族 exponential family of distributions 180
事象の独立性 independence of events 84
識別問題 identification problem 444
射影作用素 projection operator
直交―― orthogonal 9, 44
定義 44
陽表現 explicit expressions 44, 45
主成分 principal components 534
収束 convergence
確率変数の―― of random variables
r 次平均―― r-th mean 139
確率―― in probability 102
確率 1 での―― with probability one 102
2 次平均―― quadratic maen 102
――についての定理 115, 139
ほとんど確実な(強)――
almost sure (strong) 102
法則―― in law 109
関数の―― of function 127
積分の―― of integral
上限収束定理 dominated convergence theorem 128
上限収束定理（プラットの拡張） 129
単調収束定理 monotone convergence theorem 127
ファトゥーの補助定理 128
分布関数の―― of distribution functions
拡張したヘリー＝ブレイの定理 111
弱コンパクト性 weak compactness 109
積率に関する定理
theorem involving moments 113
定義 109
ヘリーの補助定理 109
ヘリー＝ブレイの定理 110
ポリヤの定理 112
法則―― in law 109
密度関数の――（シェッフェ）
of densities (Scheffe) 117
十分統計量 sufficient statistic
完備―― completeness 294
推定における利用（ラオ，ブラックウエル）
use in estimation (Rao, Blackwell) 294
定義 123
フズルバザールの予想
Huzurbazar's conjecture 125
分解定理 factorization theorem 124
有界完備―― bounded completeness 414
十分部分集合体 sufficient sub-field 133
順序統計量 order statistics 196, 382
条件つき確率 conditional probability 83
条件つき期待値
conditional expectation 89, 92, 133
条件つき分散 conditional variance 90
条件つき密度 conditional density 85, 92
情報の尺度（フィッシャー）
information measure (Fisher) 301-303
信頼集合と信頼区間
confidence sets and intervals 426
スターリングの近似
Stirling's approximation 56
スティルチェス積分 Stieltjes integral 125
推測確率 fiducial probability 310
推定 estimation
一致性（確率の）
consistency (probability) 315
一致性（フィッシャー） (Fisher) 316
一致性（方法の） (Method) 316
一般推定問題における―― 305
CAN 推定量 CAN estimator 318
CUAN 推定量 CUAN estimator 350
カルバック＝ライブラーの分離度
Kullback-Leibler separator 323
偏りのある推定量 biased estimator 304

経験ベイズ推定量（ロビンス） empiribcal Bayes estimator (Robbins) 308
継続の法則 law of succession 308
最小カイ 2 乗 minimum chi-square 322
最小 2 乗――→線形推定論，ガウス＝マルコフを見よ
最小分散法 minimum variance 290-301
最尤――→最大尤度を見よ
情報行列 information matrix 303
信頼集合 confidence set 306, 425
推測確率を用いる――（フィッシャー）
fiducial inference (Fisher) 310
推定量の併合 pooling of estimator 355
積率法 method of moments 321
同程度に不明とみなす法則
law of equal ignorance 307
得点法 method of scoring 336-338
BAN 推定量 BAN estimator 319
標本平均の非許容性（スタイン） inadmissibility of sample mean (Stein) 314
分散であらわされた情報の限界（クラメール，ラオ） information limit to variance (Cramer, Rao) 296-298
ヘリンガーの距離による――
Hellinger distance 323
ベイズ推定量 Bayes' estimator 307
ホールデンの不一致度による――
Haldane's discrepancy 323
ミニマックス法 minimax 311
有効性（1 次の，ラオ）
efficiency (first oredr, Rao) 319
有効性（漸近分散について）
(asymptotic variance) 317
有効性（2 次の，ラオ）
(second order, Rao) 323
有効性（フィッシャー） (Fisher) 319
正規分布 normal distribution
円周正規分布 circular normal 161, 164
積率母関数
moment generating function 94
特性関数 characteristic function 95, 149
正規分布（特徴づけ） (characterization)
クラメールの定理 151
最大エントロピー maximum entropy 151
ダルモア＝スキトヴィッチの定理 199
他の定理 other theorems 199
ド・モアブルの定理 150
ハーゲンの仮説 Hagen's hypothesis 150
ハーシェルの仮説
Herschell's hypothesis 147
ポリヤの定理 197
マックスウエルの仮説
Maxwell's hypothesis 149
成長モデル（ポソフ，ロイ）
growth model (Pothoff, Roy) 507
積率(モーメント)母関数
moment generating function
正規分布 normal 95
定義 94
2 項分布 binomial 95
ポアソン分布 Poisson 95
0-1 法則 zero-one law 139
線形推定論（統一理論） linear estimation
IPM 法 IPM method 271
一般ガウス＝マルコフ模型 270
一般逆行列に関する基本的補助定理
basic lemma on g-inverse 267
線形仮説の検定
test of linear hypotheses 273
分散共分散 variances and covariances 271
モデルが意味をもつことの検定
test for consistency 270
ULS 法 ULS method 273
相関 correlation→回帰を見よ

タ　行

ダンデカーの連続性修正
Dandekar's continuity correction 378
多項分布 multinomial distribution
推定 estimation 326-333
定義 326
多変量正規分布
multivariate normal distribution
対称―― symmetric 182
定義 170, 468, 472
特性関数 characteristic function 170, 470
特徴づけ characterization 475, 476, 481
2 変量正規 bivariate normal 186
――の性質 properties of 469-475
母数の推定 estimation of parameters 478
密度（正則な分散共分散行列をもつ場合）
density (nonsingular dispersion)

170, 179, 468, 477
密度（特異な分散共分散行列をもつ場合）
(singular dispersion) 477
多変量積率（モーメント）
multivariate moments 99
対角和（トレース）最小問題
minimum trace problems 61
大数の法則 law of large numbers
コルモゴロフの定理 (SLLN) 106, 107
コルモゴロフの定理 (WLLN) 137
その他の定理 137
チェビシェフの定理 (WLLN) 105
定義 (WLLN, SLLN) 104
ヒンチンの定理 (WLLN) 105, 137
逐次解析 sequential analysis
A. S. N. 関数 A. S. N. function 439
管理図（の動作特性）
control chart (O. C.) 461
基本等式 fundamental identity 437
検出力が1の検定 test with power one 441
推定（シュタイン） estimation (Stein) 415
逐次確率比検定
(S. P. R. T.)（ワルド）(Wald) 430-433
逐次確率比検定のうちきり
truncation of S. P. R. T. 461
逐次確率比検定の効率
efficiency S. P. R. T. 433
動作特性関数 O. C. function 438
中心極限定理 C. L. T. central limit theorems
多変量の場合 multivariate case 121, 138
不変性原理 invariance principle 139
リャプノフ 120
リンドバーグ＝フェラー 120
リンドバーグ＝レヴィ 119
重複対数の法則 law of iterated logarithms 122, 442
t 検定（スチューデント） t-test
158, 169, 183, 217, 222, 416
定理（発見者の名のついた） theorems
クラメール（正規） Cramer 151
グリベンコ Glivenko 383
ケーレイ＝ハミルトン（行列）
Cayley-Hamilton 42
コルモゴロフ (SLLN) Kolmogorov
106, 107
コルモゴロフ（確率変数の和） 122
コルモゴロフ (WLLN) 137
コルモゴロフ（D_n の分布関数） 384
コルモゴロフ（分布の一致） 100
坂元＝クレーグ Sakamoto-Craig 193
シェッフェ Scheffe 117
ストルムの分離——
Strumian separation 60
スミルノフ Smirnov 384
ダルモア＝スキトヴィッチ
Darmois-Skitovic 199, 476
チェビシェフ (WLLN) Chebyshev 105
ド・モアブル De Moivre 150
ヒンチン (WLLN) Khinchin 105, 137
フィッシャー＝コクラン
Fisher-Cochran 171
フィーラー Fieller 283
フビニ Fubini 129
フロベニウス（正行列） Frobenius 43
ヘリー＝ブレイ Helly-Bray 110
ヘリー＝ブレイ（拡張） 111
ベイズ Bayes 305
ペロン（正行列） Perron 43
ボホナー（特性関数の非負定符号）
Bochner 133
ポアンカレの分離——
Poincare separation 60
ポリヤ（正規分布） Polya 197
ポリヤ（分布関数の収束） 111
ラドン＝ニコディム Radon-Nikodym 129
リャプーノフ (C. L. T.) Liapunov 120
リンドバーグ＝フェラー (C. L. T.)
Lindberg-Feller 120
リンドバーグ＝レヴィ (C. L. T.)
Lindberg-Levy 119
ルベーグの上限収束—— Lebesgue
dominated convergence 128
ルベーグの単調収束—— Lebesgue
monotone convergence 127
統計量と部分集合体
statistics and subfields 131
統計量の変換 transformation of statistics
相関係数 ($\tanh^{-1}$)
correlation coefficient 392
2項割合 ($\sin^{-1}$) binomial proportion 389
ポアソン変数 ($\sqrt{\ }$) Poisson variable 388
同程度に不明とみなす法則
law of equal ignorance 307
特異値分解 singular value decomposition 40
特性関数 characteristic function
キュミュラント cumulants 94
積率（モーメント） moment 87, 94
定義 92
——の収束 convergence of 111
——の性質 properties of 93
——の展開 expansion of 93
——の非負定符号性
non-negative definiteness of 133
——の例 134
反転定理 inversion theorem 97
分布の——
ガンマ分布 gamma 141
コーシー分布 Cauchy 141
3角形分布 triangular 141
正規分布 normal 95
双曲線分布 hyperbolic 141
多変量正規分布
multivariate normal 170, 470
退化した分布 degenerate 141
長方形分布 rectangular 141
2項分布 binomial 95
ベータ分布 beta 141
ポアソン分布 Poisson 95
ラプラス分布 Laplace 141
ボホナーの定理 Bochner's theorem 133
連続性定理 continuity theorem 111
得点による検定（ラオ） scoring test 380, 381
凸集合 convex sets
定義 definition 48
分離定理 separation theorems 49

ナ　行

並べかえによって生ずる分布
permutation distribution 455
2次形式 quadratic form
——の極値問題 extremum problems 56-61
——の分布 distribution of 171-179
——の分類 classification 32
——の変換 transformation 33
1つの——の標準化
reduction of one form 19, 37
2つの——の標準化
reduction of two forms 38
2次の有効性（ラオ）
second order efficiency 323
2変量正規分布
bivariate normal distribution
回帰係数の分布 distribution of regression coefficient 191
性質 properties 186
相関係数の分布 distribution of
correlation coefficient 190, 191
特性関数 characteristic function 187
分散共分散の分布 distribution of
variances and covariances 189
平均値の分布 distribution of means 189
密度 density 186
ノンパラメトリック推測
nonparametric inference
ウイルコクソン検定 Wilcoxon test 453
頑健性 robustness 450
並べかえによって生ずる分布
permutation distribution 455
ファン・デル・ヴェルデン検定
Van der Waerden test 453
フィッシャー＝イェーツ検定
Fisher-Yates test 453
符号検定 sign test 452
分布によらない方法
distribution free method 451
無作為化の原理
randomization principle 453
ワルド＝ウォルフォヴィッツ検定
Wald-Wolfowitz test 453

ハ　行

判別分析 discriminatory analysis
判別関数の検定 test on discriminant
functions 515, 542
判別関数の十分性 sufficiency of
discriminant function 523
判別得点 discriminant scores 518
フィッシャーの判別関数
Fisher's discriminant function 513
複合仮説の判別 discrimination of
composite hypotheses 524
標本分布（正確な）
sampling distributions (exact)
ウィシャート分布 Wishart distribution 540

カイ2乗（中心） chi-square (central) 168
カイ2乗（非心） (noncentral) 168
回帰（2変量）の―― regression 191
級内相関（0のとき）の――
correlation, intraclass (null) 184
級内相関（0でないとき）の――
(non-null) 184
重相関（0のとき）の――
correlation, multiple 249
重相関（0でないとき）の―― 541
スチューデントの t（中心） Student's t 158
スチューデントの t（非心） 159
正規変数の2乗の――
square of normal variable 152
（積率）相関（$\rho=0$ のときの）の――
correlation, product moment 190
（積率）相関（$\rho\neq0$ のとき）の―― 191
2次形式の―― quadratic forms 171-179
p_1（ピアソン） (Pearson) 157
標本平均と標本分散共分散の――
sample mean and dispersion 486
標本平均と標本分散の――
sample mean and variance 169
フィッシャーの F（中心） Fisher's F 156
フィッシャーの F（非心） 198
フィッシャーの z Fisher's z 156
ホテリングの T^2（中心） Hotteling's T^2 491
ホテリングの T^2（非心） 491
マハラノビスの D^2 Mahalanobis' D^2 491
U 統計量（付加情報，ラオ） U-statistic 502
標本分布（漸近的） (asymptotic)
一様性の H 検定
H-test for homogeneity 355
ウイルクスの Λ Wilk's Λ 503
コルモゴロフ＝スミルノフ検定
Kolmogorov-Smirnov test 384
適合度の検定のための χ^2 分布（ピアソン）
χ^2 for goodness of fit 357
度数の1次関数
linear functions of frequencies 349
度数の2次関数
quadratic function of frequencies 350
統計量の関数 functions of statistics
352-355
得点による検定（ラオ） scoring test 380
特定のいくつかの級での偏差
deviation in specified cells 363
1つの級での偏差
deviation in a single cell 359
標準誤差 standard errors 353-354
分位点 fractiles 385-387
分割表における χ^2 分布
χ^2 for contingency tables 367, 369-375
尤度比検定（ネイマン＝ピアソン） likelihood
ratio test (Neyman-Pearson) 380
ワルドの大標本検定
Wald's large sample tests 380-382
フィッシャーの F Fisher's F 156
フィッシャーの z Fisher's z 156
フズルバザールの予想（十分性）
Huzurbazar's conjecture 125
フレイザー＝スプロットの公式
Fraser-Sprott equation 190
不等式（確率的） inequality (probability)
イェンセン Jensen 140
カンテリ Cantelli 136
カンプ＝メイデル Camp-Meidell 136
コルモゴロフ Kolmogorov 135
チェビシェフ Chebyschev 89
チェビシェフ型 Chebyschev type 136
ハイエク＝レニュイ Hajek-Renyi 106, 135
ピーク Peek 136
ヘルダー Hölder 140
ベルジュ Berge 137
ミンコフスキ Minkowski 140
不等式（代数的） inequality (algebraic)
アダマール Hadamard 53
イェンセン Jensen 55
カントロビッチ Kantorovich 69
コーシー＝シュヴァルツ Cauchy-Schwarz 51
シルベスターの法則 Sylvester's law 29
情報理論における――
information theory 55
積率に関する―― moments 53
フロベニウス Frobenius 29
ヘルダー Hölder 52
ベッセル Bessel 10
ミンコフスキ Minkowski 53
ラグランジュの等式 Lagrange identity 52
不変性 invariance
確率測度のための――原理 principle for
probability measures 130

CLT のための――原理 139
推定における――原理
principle in estimation 314
統計的手順のための――原理
statistical procedure 131
部分集合体 subfield 81, 131, 133
分割表 contingency tables
イェーツの連続性修正
Yates' continuity correction 377, 378
属性の独立性の検定 test for
independence of attributes 369-378
ダンデカーの連続性修正
Dandekar's continuity correction 378
分散共分散行列 dispersion matrix 100, 406
分散共分散成分
variance and covariance components
一般模型 general model 276
2元データの―― two way data 236
MINQUE 理論 minque theory 276
分散共分散分析 analysis of dispersion
ウイルクスの Λ 分布
distribution of Wilks' Λ 502, 503
ガウス＝マルコフ（多変量）
Gauss-Markoff (multivariate) 491
構造関係の検定
test for structural relationship 511
次元数の検定 dimensionality 503
成長モデル（多変量） growth model 507
線形仮説の検定 linear hypotheses 495-499
定義 definition 496
付加情報の検定
test for additional information 499
分散分析 analysis of varaince
1元分類 one-way classification 224
一般ガウス＝マルコフ模型 270
F 検定（フィッシャー） F-test 218, 219
回帰 → 回帰と関連性を見よ
ガウス＝マルコフ模型 201
区間推定 interval estimation 218, 426
共分散による調整
covariance adjustment 263
t 検定（スチューデント，ゴセット）
t-test (Student, Gosset) 217, 223
同時信頼区間 simultaneons confidence
intervals 219
2元分類，分散成分 two-way classifi-
cation, variance components 236
繰返し数が等しいとき
with equal numbers 226, 231
繰返し数が等くしないとき
with unequal numbers 232
比の信頼区間
confidence interval for ratio 221
非加法性の検定 test for nonadditivity
一般 general 230
テューキーの―― Tukey's 228, 229
MINQUE（分散成分） 61-63, 237, 276
分布関数 distribution function
条件つき―― conditional 85
定義 79
――の性質 properties 79-81
――のたたみこみ convolution 96
――の分解 decomposition 80
分布族の内在的精度（フィッシャー）
intrinsic accuracy (Fisher) 303
分布によらない方法
distribution free methods 451
ベイズ決定方式 Bayes decision rule 446
ベクトル空間 vector space
基底（ハーメル） basis (Hamel) 4
グラム＝シュミットの直交化
Gram-Schmidt orthogonalization 9
正規直交基底 orthonormal basis 9
直交基底 orthogonal basis 9
直和 direct sum 11, 38
定義 3
内積 inner product 8, 68
部分空間 subspace 3
スタイニッツの取替問題
Steinitz replacement result 13
変換（1次） transformation (linear) 22
変数の変換 transformation of variables 146
ホテリングの T^2 Hotelling's T^2 490
ボレル集合体（σ 集合体） Borel field 77
補助情報 ancillary information 457
補助定理（発見者の名のついた） lemmas
クロネッカー Kronecker 139
シュール Schur 72
ネイマン＝ピアソン Neyman-Pearson 405
ファトゥー Fatou 128
ファルカス Farkas 50
ヘリー Helly 109

ボレル＝カンテリ　Borel-Cantelli　130

マ　行

マハラノビスの D^2　Mahalanobis D^2　490
MINQUE 理論（分散成分）　minque theory (variance components)　276
密度　density
ウィシャート分布　Wishart　539
f 分布（中心）　f (central)　154
f 分布（非心）　f (noncentral)　160, 198
F 分布（中心），フィッシャー　156
F 分布（非心），フィッシャー　198
円周正規分布　circular normal　163, 164
カイ2乗分布（中心）　chi-square　155, 168
カイ2乗分布（非心）　168
ガンマ分布　gamma　141, 153
回帰係数　regression coefficient　191
幾何分布　geometric　82
逆正弦分布　arc sine　154
コーシー分布　Cauchy　141, 157
3角形分布　triangular　141
指数型分布族　exponential family　180
寿命試験のモデル分布
life testing model　196
重相関（中心）　multiple correlation　249
重相関（非心）　541
順序統計量　order statistics　196
スチューデントの t 分布（中心）
Student's t　158, 169, 183
スチューデントの t 分布（非心）　159
z（フィッシャー）分布　(Fisher)　156
正規分布　normal　82
双曲余弦分布　hyperbolic cosine　141
相関係数分布（$\rho=0$）
correlation coefficient (null)　190
相関係数分布（$\rho\neq0$）(non-null)　191
多項分布　multinomial　326
多変量正規分布　multivariate normal　171, 179, 477
対称正規分布　symmetric normal　182
対数正規分布　lognormal　195
対数分布　logarithmic　82
退化した分布　degenerate　141
長方形分布　rectangular　141
超幾何分布　hypergeometric　82
ディリクレ分布　Dirichlet　197
2項分布　binomial　82, 95, 145
2変量コーシー分布　bivariate Cauchy　157
2変量正規分布　bivariate normal　186
パレート分布　Pareto　195
ピアソン系分布　Pearsonian curves　195
ピアソンの p_2 分布　Pearson p_2　157
フォン・ミーゼス分布　Von Mises　164
負の2項分布　negative binomial　82
ベータ分布（中心）beta　141, 154, 156
ベータ分布（非心）　159, 198
ベッセル関数分布　Bessel function　199
ホテリングの T^2 分布（中心）
Hotelling's T^2　490
ホテリングの T^2 分布（非心）　490
ポアソン分布　Poisson　82, 95
方向に関するデータ分布
directional data　163
マケハムの法則分布　Makeham's law　196
マックスウエル分布　Maxwell　161
巻きこみ分布　wrapped-up distributions　165
待ち時間分布　waiting time　154
ラプラス分布　Laplace　141, 195
粒子の平衡状態分布
equilibrium state of particles　160
レイリー分布　Rayleigh　153
ロジスティック分布　logistic　195
ワイブル分布　Weibull　196
無作為化の原理　randomization principle　453
無限分解可能な分布
infinitely divisible distribution　140
積率問題（モーメント問題）
moment problem　98

ヤ，ラ，ワ　行

尤度　likelihood　323
尤度比（単調）
likelihood ratio (monotone)　418
尤度比検定（ネイマン＝ピアソン）
likelihood ratio test　380, 381
ルベーグ＝スティルチェス積分
Lebesgue-Stieltjes integral　125
ルベーグ積分　Lesbesgue integral　126
ワルドの大標本検定
Wald's large sample tests　380, 381
ワルドの逐次確率比検定
Wald's SPRT　430

訳者あとがき

この本は，1973年に刊行された，C. R. Rao の“*Linear Statistical Inference and Its Applications*”の第2版の訳書である．

C. R. ラオは，統計学の分野で，いまや世界で最高の権威者の1人であり，彼の著述は，理路整然として簡潔であるという点で定評がある．かれの数学の力と独創性は見事なものであり，その証明がエレガントであることは特に有名である．完全な解がまだ得られていないような難かしい問題について，現在入手できる知識を整理して人に紹介する能力においては，ラオの右に出る人はいないであろう．

そのような人物がまとめた本であるから，1965年にこの第1版が刊行されると，わが国でも，いくつかのグループがこの本をテキストとして，1年ないし2年にわたり輪読会を開いたと聞く．その結果，統計学の理論および応用に関心をもつ多くの人々に大きな感銘を与え，その知識を拡充し深めるのに役立ったと考えられる．

この本の最初の2章は，それだけで完結した入門コースを構成しており，第1章は　ベクトル空間と行列の理論，第2章は　確率論とその応用，についての必要な基本的概念のすべてを論述している．この確固たる数学的基礎のあとに，連続型確率模型，最小2乗法の理論と分散分析，推定の基準と方法，大標本理論と方法，統計的推測の理論，多変量解析，という重要な章がつづく．それらの各章では理論とともに，応用分野での広汎な話題が取り上げられている．

第2版で新しく取り入れられた論題は，線形推定に関して分散共分散行列が特異な場合を含む，著者の新しい統一理論，分散成分推定のための MINQUE 理論，一般逆行列とその応用，正行列，ベクトルと行列の微分，制限つき回帰問題，偏りのある推定，検出力1の逐次抜取，判別分析の最近の発展，成長モデル，補助情報の利用，統計量と部分集合体の概念，頑健性，などである．これらの新しい論題は，大標本理論，多変量解析，線形模型，推定論，推測理論の書き改められた叙述のなかにうまく統合されている．そのうえ，重要な概念

を例示するのに，現実世界への応用と関連づけて，沢山のなまの例題が取り扱われている．

このように見てくると，この本が教科書としていかにすぐれたものであるかが明らかであり，現在の統計理論と統計的方法の真髄を包含し，その内容においても叙述においても，これに匹敵するものを他に見出すことはできない．したがって，統計学を勉強しようとする学生や，いろいろの分野での統計専門家は，この本をバイブルとして，その富豊な内容から有用な知識を得て，学習，研究，応用に役立ててほしいと考える．

C. R. ラオは，1960年以来，数回にわたる来日のたびごとに，日本の統計学者や応用分野の人々に，なんらかの講義またはセミナーを行っている．また，わが国からは，ほとんど毎年，ラオの所属するインド統計研究所に招かれて，講義やセミナーを担当する人が渡印し，その数は十指に余るようになった．このように，日・印のあいだの統計家の交流は，他の国と比べて特に密接である．1973年に来日したとき，ラオから「この本はすでに10数ヶ国語に翻訳されている．どうして日本語訳が出ないのか」と言われたのが契機で，筆者はこの訳業に参加することになった．しかし，これだけの大部の本を1人で訳すことは到底不可能なので，数人の共訳者をお願いし，月に一度の割で会合しながら，訳語の統一，理解困難な箇所の討論などを行ってきた．訳の分担は，次のとおりである．

第1章　奥野	第5章　鷲尾・篠崎
第2章　矢島・古河	第6章　長田
第3章　長田	第7章　矢島・古河
第4章　鷲尾	第8章　広崎

訳出のあと，原稿またはゲラ刷での調整は，奥野が行った．その際，理論的な面では，日本 IBM の渋谷政昭氏に，いろいろ御教示を得た．ここに記して厚く謝意を表する．また，東大工学部計数工学科の次の諸君には，原著の数値の誤り，誤植の訂正までを含めて，奥野の仕事を助けて頂いた．その分担は次のとおりでここに記してこれらの諸君にも謝意を表したい．

中島信和（1章），西尾敦（2章），三輪哲久（3章），福村聡（4章），

大橋靖雄（5，8章），柴田義貞（6章），藤野和建（7章）．

かれらの見つけた誤りは，すでにラオにも知らせて承認を得，謝辞をもらっている．それらの誤りは訳書では断らずに訂正した．

本年初頭になって，ラオから重要な訂正が送られてきた．それは，次の箇所である．

	原　著	訳　書
①	43頁8行目～13行目	40頁12行目～25行目
②	45頁8行目～11行目	42頁16行目～27行目
③	63頁18行目～64頁8行目	59頁26行目～60頁20行目
④	323頁17行目～324頁4行目	296頁2行目～16行目
⑤	484頁3行目の文章の後に1行追加	438頁18行目の1行
⑥	512頁の Bahadur (1960) の次に追加	463頁7行目～8行目

これらの箇所はラオの訂正済みのものを訳しなおして入れた．そのほか，細かい訂正を121箇所も指摘してきた．その大部分は当方で気付いていたものであるが，もちろん，こちらの見おとしもあった．それらはすべて訳書では断りなしに訂正しておいたので，原書との食違いを発見された読者は，そこが訂正されたものであるかどうかを判断されたい．

そのほかに，日本語訳では注記をしたほうが良いと思われる箇所には〔　〕づきで訳者注を入れた．

この訳業が4年もかかったのは，主として筆者の怠慢に依る．早くに仕事を済まされた共訳者各位にお詫びするとともに，その間よく忍耐して訳の完成を待ち，かつ，初校・再校での大幅な赤字入れにも寛容さを示し，また，索引の作成にも全面的に協力していただいた，東京図書の須藤静雄氏にはなんとお詫びし，謝意を表してよいかわからない次第である．

1977年10月

奥野忠一

C. R. ラオのプロファイル

1. 出　生　1920年9月10日，Karnataka 州 Hadagali 生
2. 学　歴　1940　Andhra 大学数学科1番で卒業
　　1943　Calcutta 大学統計学科1番，金賞で卒業
　　1948　論文 "Statistical Problems of Biological Classification" で Cambridge 大学の Ph. D
　　1965　統計学に関する著作で Cambridge 大学の Sc. D
3. 職　歴　1949—63　インド統計研究所，Research and Training School (RTS) 教授兼 Head
　　1964—72　同所 RTS, Director
　　1972—76　インド統計研究所，Secretary and Director
　　1976—　同所，Jawaharlal Nehru Professor
　（海外）1946—48　Research Fellow, Cambridge Univ., UK
　　1953—54　Visiting Research Professor, Illinois Univ.
　　1954（夏）Visiting Professor, Univ. of California, Berkeley
　　1963—64　Senior Research Fellow, Johns Hopkins Univ.
　　1964（夏）Visiting Professor, Stanford Univ.
　　1973　Patton Foundation Professor, Indiana Univ.

そのほか UK, USA, USSR, Canada, Australia, 日本およびヨーロッパ各国の大学および研究所に招かれて講義．1977年から International Statistical Institute の President で，77年12月にニューデリーで開かれる第41回総会を主宰する．

4. 研究業績

推定論　クラメール＝ラオの不等式，ラオ＝ブラックウェル化，フィッシャー＝ラオの定理，線形推定の統一理論（ガウス＝マルコフ模型），分散成分の MINQUE 理論

組合せ論　ラオの直交配列，ハミング＝ラオ限界

特徴づけ問題　カガン＝リニック＝ラオの定理，ラオ＝ルービンの定理

多変量解析　ラオのU検定，分散共分散分析とクラスター分析，連結横断面の研究

行列代数　一般逆行列，補集合のない部分空間における射影演算子

訳者紹介

奥野忠一（おくの ただかず）
1944年　東京大学理学部数学科卒業
現　在　東京理科大学経営工学科教授

長田　洋（おさだ ひろし）
1970年　東京大学工学部計数工学科卒業
現　在　旭化成工業株式会社勤務

篠崎信雄（しのざき のぶお）
1969年　慶応義塾大学工学部管理工学科卒業
現　在　慶応義塾大学総合政策学部教授

広崎昭太（ひろさき しょうた）
1950年　宮崎農業専門学校卒業
　　国立公害研究所環境情報部長
1991年　没

古河陽子（ふるかわ ようこ）
1966年　日本大学文理学部卒業
現　在　日本科学技術研修所情報システム部勤務

矢島敬二（やじま けいじ）
1954年　東京都立大学理学部数学科卒業
現　在　東京理科大学工学部教授

鷲尾泰俊（わしお やすとし）
1953年　九州大学理学部数学科卒業
現　在　慶応義塾大学理工学部教授

統計的推測とその応用

1977年11月25日　第1刷発行　Printed in Japan
1992年4月30日　新装第1刷発行

検印省略

著　者　C.　R.　ラ　オ
訳　者　奥　野　忠　一ほか
発行所　東京図書株式会社
〒112　東京都文京区水道2—5—22
振替東京4-13803 電話03(3814)7818～9

ISBN 4-489-00381-1